Module 3: Random Variables and Vectors

Ruben Zamar
Department of Statistics
UBC

September 27, 2016

Random Variable

- Let $(\Omega, \mathcal{F}, P)$ be a probability space.

- Definition of random variable:

$$X \quad : \quad \Omega \rightarrow \mathcal{R}$$

$$X^{-1}(B) \quad \in \quad \mathcal{F}, \qquad \text{for all} \quad B \in \mathcal{B}(R) \qquad (\text{measurable function})$$

- Induced probability:

$$P_X(B) = P\left(X^{-1}(B)\right), \qquad B \in \mathcal{B}(R).$$

Distribution Function

- **Definition:**

$$
\begin{aligned}
F(x) &= P(X \leq x) \\
&= P\left(X^{-1}(-\infty, x]\right) \\
&= P_X\left((-\infty, x]\right)
\end{aligned}
$$

- **Example.**

$$
\begin{aligned}
(\Omega, \mathcal{F}, P) &= \left((0,1), \mathcal{B}\left((0,1)\right), P\right), \\
P\left((a,b)\right) &= b-a, \text{ for all } 1 > b > a > 0.
\end{aligned}
$$

$$X(w) = w^2,$$

$$
\begin{aligned}
F_X(x) &= P(X \leq x) = P\left(\left\{w : w^2 < x\right\}\right) \\
&= P\left(\left(0, \sqrt{x}\right)\right) = \sqrt{x}, \quad x \in (0,1).
\end{aligned}
$$

Pi-Class

- **Definition of π-class:** a collection of sets which is **closed under finite intersections**.

- **Result:** Let P, Q be two probability functions such that $P(A) = Q(A)$ for all $A \in \mathcal{C}$, where $\mathcal{C}$ is a π-class. Then $P(A) = Q(A)$ for all $A \in \mathcal{F}(\mathcal{C})$, where $\mathcal{F}(\mathcal{C})$ is the σ-field generated by $\mathcal{C}$.

- Since
$$\mathcal{C} = \{(-\infty, x] : \text{for some } x \in R\}$$
is a π-**class**, $F(x)$ completely determines the induced probability function, $P_X(B)$.

- Recall that
$$\text{Borel Sets} = \mathcal{F}(\mathcal{C})$$

Properties of F

- $P(a < X \leq b) = F(b) - F(a)$

Proof.

$$\begin{aligned}
(X \leq b) &= (X \leq a) \cup (a < X \leq b) \\
&\Longrightarrow P(X \leq b) = P(X \leq a) + P(a < X \leq b) \\
&\Longrightarrow F(b) = F(a) + P(a < X \leq b).
\end{aligned}$$

- $\lim_{x \to -\infty} F(x) = \lim_{x \to -\infty} P(X \leq x) = P(\phi) = 0$

Poof.

$$(X \leq x) \quad \downarrow \quad \cap (X \leq x) = \phi, \quad \text{as} \quad x \to -\infty.$$

$$\Rightarrow \quad \lim_{x \to -\infty} P(X \leq x) = P(\phi) = 0$$

[Using continuity property of P]

- $\lim_{x\to\infty} F(x) = \lim_{x\to\infty} P(X \le x) = P(\Omega) = 1$

Proof.

$$(X \le x) \quad \uparrow \quad \cup (X \le x) = \Omega, \text{ as } x \to \infty.$$

$$\Rightarrow \quad \lim_{x\to\infty} P(X \le x) = P(\Omega) = 1$$

- $F(x)$ is right-continuous:

$$\lim_{x \downarrow x_0} F(x) = \lim_{x \downarrow x_0} P(X \leq x) = F(x_0)$$

Proof.

$$(X \leq x) \downarrow \cap_{x > x_0} (X \leq x) = (X \leq x_0), \quad \text{as } x \downarrow x_0.$$

Then use continuity property of P.

- $P(X = x) = F(x) - F(x^-)$

Proof.

$$\{x\} = \cap_{\epsilon > 0}(x - \epsilon, x + \epsilon],$$

$$\Longrightarrow \quad P(X = x) = \lim_{\epsilon \to 0} P(x - \epsilon < X \leq x + \epsilon)$$

$$= \lim_{\epsilon \to 0} F(x + \epsilon) - \lim_{\epsilon \to 0} F(x - \epsilon)$$

$$= F(x) - F\left(x^-\right)$$

- $F(x)$ is non decreasing and has at most a countable number of jumps.

  Proof: Non decreasing:

$$
\begin{aligned}
x_1 &< x_2 \\
&\Rightarrow \{X \leq x_1\} \subset \{X \leq x_2\} \\
&\Rightarrow F(x_1) \leq F(x_2).
\end{aligned}
$$

- **Countable number of jump discontinuities:** Let

$$A_n = \left\{x : F(x) - F\left(x^-\right) \geq 1/n\right\}$$

Clearly

$$\#A_n \leq n \quad \text{(why?)}.$$

Hence

$$\left\{x : F(x) - F\left(x^-\right) > 0\right\} = \cup_{n\geq 1} A_n$$

is a countable set (finite or infinite).

Discrete Random Variables

- The range of X is finite or countable:

$$X(\Omega) = \{x_i\}_{i \in I}$$

$$I = \{1, 2, ... n\} \quad \text{or} \quad I = \mathcal{N}^+ \text{ (positive integer numbers)}$$

- Assume (w.l.g.) that $x_1 < x_2 < x_3 < \cdots$

- For most discrete r.v.'s

$$X(\Omega) \subset \mathcal{N} \qquad \text{(integer numbers)}$$

Probability Mass Function

- **Definition:**

$$f\left(x_i\right) \quad = \quad P\left(X = x_i\right)$$

- **Properties:**

(1) $f\left(x_i\right) \geq 0 \quad ,$

(2) $\sum_{i \in I} f\left(x_i\right) = 1$ and

(3) $f\left(x_i\right) = F\left(x_i\right) - F\left(x_{i-1}\right).$

Continuous Random Variable

- $F(x)$ is determined by a **probability density function** $f(x)$**:**

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$$

- The density function, $f(x)$, satisfies 2 properties

(1) $f(x) \geq 0$ for all x, and

(2) $\int_{-\infty}^{\infty} f(x)\, dx = 1.$

- By the Fundamental Theorem of Calculus, $F(x)$ is differentiable and

$$f(x) = F'(x)$$

Expected Value

Let X be a random variable with density function $f(x)$. Then

$$E(X) = \sum_{i \in I} x_i f(x_i) \qquad \text{(discrete case)}$$

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx \qquad \text{(continuous case)}$$

provided

$$\sum_{i \in I} |x_i| f(x_i) < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$$

More generally,

$$E(g(X)) = \sum_{i \in I} g(x_i) f(x_i) \qquad \text{(discrete case)}$$

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) \, dx \qquad \text{(continuous case)}$$

provided

$$\sum_{i \in I} |g(x_i)| f(x_i) < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} |g(x)| f(x) \, dx < \infty$$

EXAMPLES OF RANDOM VARIABLES

Bernoulli Random Variable

Notation: $X \sim \text{Bernoulli}(p)$

$$X(w) = I_A(w) = 1, 0$$

depending on whether $w \in A$ or not. Let

$$p = P(A)$$

Then

$$P(X = x) = p^x (1 - x)^{1-x}, \quad x = 0, 1$$

In this case

$$\begin{aligned} E\left(X^k\right) &= 0 \times (1 - p) + 1 \times p = p, \quad k = 1, 2, \ldots \\ Var(X) &= E\left(X^2\right) - (E(X))^2 = p - p^2 = p(1 - p) \end{aligned}$$

Binomial Random Variable

Notation: $X \sim B(n, p)$

$$f(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, ..., n$$

$$\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = [p + (1-p)]^n = 1$$

$$
\begin{aligned}
E(X) &= \sum_{x=0}^{n} x \begin{pmatrix} n \\ x \end{pmatrix} p^{x} (1-p)^{n-x} \\
&= \sum_{x=1}^{n} \frac{n!}{(x-1)!\,(n-x)!} p^{x} (1-p)^{n-x} \\
&= np \sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\,(n-x)!} p^{x-1} (1-p)^{n-x} = np
\end{aligned}
$$

$$
\begin{aligned}
E\left(X\left(X-1\right)\right) &= \sum_{x=0}^{n} x\left(x-1\right)\left(\begin{array}{c} n \\ \\ x \end{array}\right) p^{x}\left(1-p\right)^{n-x} \\
&= \sum_{x=2}^{n} \frac{n!}{\left(x-2\right)!\left(n-x\right)!} p^{x}\left(1-x\right)^{n-x} \\
&= n\left(n-1\right)p^{2} \sum_{x=2}^{n} \frac{\left(n-2\right)!}{\left(x-2\right)!\left(n-x\right)!} p^{x-2}\left(1-p\right)^{n-x} \\
&= n\left(n-1\right)p^{2}
\end{aligned}
$$

$$
\begin{aligned}
n(n-1)p^2 &= E(X(X-1)) = E\left(X^2\right) - E(X) \\
&= E\left(X^2\right) - np
\end{aligned}
$$

$$
\Rightarrow \quad E\left(X^2\right) = n(n-1)p^2 + np
$$

$$
\begin{aligned}
Var(X) &= n(n-1)p^2 + np - n^2p^2 \\
&= np(1-p)
\end{aligned}
$$

In R you can use the functions:

rbinom(m,n,p)	generates m random variables
dbinom(0:m,n,p)	probabilities of 0,1,...,m
pbinom(0:m,n,p)	cdf at 0,1,...,m
qbinom(q,n,p)	quantile function (see latter)

Geometric Random Variable

Notation: $X \sim Geom\,(p)$

$$f\,(x) = P\,(X = x) = (1-p)^x\,, \quad x = 0, 1, \ldots$$

$$\begin{aligned}
\sum_{x=0}^{n} f\,(x) &= p \sum_{x=0}^{\infty} (1-p)^x = p \frac{1}{1-(1-p)} = 1 \\
F\,(x) &= \sum_{y=0}^{x} f\,(y) = 1 - (1-p)^{x+1}\,, \quad x = 0, 1, \ldots \\
E\,(X) &= \frac{1-p}{p}, \qquad Var\,(X) = \frac{1-p}{p^2}
\end{aligned}$$

In R you can use the functions:

rnbinom(m,1,p)	generates m random variables
dnbinom(0:m,1,p)	probabilities of 0,1,...,m
pnbinom(0:m,1,p)	cdf at 0,1,...,m
qnbinom(q,1,p)	quantile function (see latter)

Note: the second parameter is the size. Variable represents number of failures before the first success. Instead of first success (size =1) could be the kth success (size = k)

The general case is called: **negative binomial.**

Poisson Random Variable

Notation: $X \sim \mathcal{P}(\lambda)$

$$f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \quad x = 0, 1, \ldots$$

$$\sum_{x=0}^{\infty} f(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x}}{x!} = e^{-\lambda} e^{\lambda} = 1$$

$$E(X) = Var(X) = \lambda$$

In R you can use the functions:

rpois(m,λ)	generates m random variables
dpois(0:m,λ)	probabilities of 0,1,...,m
ppois(0:m,λ)	cdf at 0,1,...,m
qpois(q,λ)	quantile function (see latter)

Normal Random Variable

Notation: $X \sim N\left(\mu, \sigma^2\right)$

$$f\left(x\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \qquad -\infty < \mu < \infty, \quad \sigma^2 > 0$$

To show that

$$I = \int_{-\infty}^{\infty} f\left(x\right) dx = 1$$

take $\mu = 0$, $\sigma = 1$ (wlg)

Consider

$$
\begin{aligned}
I^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} dy \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(x^2+y^2\right)} dxdy.
\end{aligned}
$$

Change of variables (polar coordinates, $r > 0$, $0 \leq \theta < 2\pi$):

$$\begin{aligned} x &= r\cos(\theta) \\ y &= r\sin(\theta) \end{aligned}$$

$$x^2 + y^2 = r^2\cos^2(\theta) + r^2\sin^2(\theta) = r^2$$

$$J = \left| \det \begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix} \right| = r\cos^2(\theta) + r\sin^2(\theta) = r$$

$$
\begin{aligned}
I^2 &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(x^2+y^2\right)} dxdy \\
&= \frac{1}{2\pi}\int_{0}^{\infty}\int_{0}^{2\pi} re^{-\frac{1}{2}r^2} d\theta dr = \int_{0}^{\infty} re^{-\frac{1}{2}r^2} dr \\
&= -e^{-\frac{1}{2}r^2}\Big|_{0}^{\infty} = 0 - (-1) = 1.
\end{aligned}
$$

Mean

$$
\begin{aligned}
E(Z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z \exp\left(-\frac{1}{2}z^2\right) dz \\
&= -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{d}{dz} \left\{ \exp\left(-\frac{1}{2}z^2\right) \right\} dz \\
&= -\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right) \Big|_{-\infty}^{\infty} = 0
\end{aligned}
$$

Variance

$$E\left(Z^2\right) = \frac{2}{\sqrt{2\pi}} \int_0^\infty z^2 \exp\left(-\frac{1}{2}z^2\right) dz$$

By part integration:

$$u = z, \qquad du = dz$$

$$dv = -z\exp\left(-\frac{z^2}{2}\right), \qquad v = \exp\left(-\frac{z^2}{2}\right)$$

$$E\left(Z^2\right) = \frac{2}{\sqrt{2\pi}} \left[z\exp\left(-\frac{z^2}{2}\right)\Bigg|_0^\infty + \int_0^\infty \exp\left(-\frac{z^2}{2}\right) dz \right]$$

$$
\begin{aligned}
E\left(Z^2\right) &= = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} \exp\left(-\frac{z^2}{2}\right) dz \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{z^2}{2}\right) dz = 1
\end{aligned}
$$

Notation: $X \sim Gamma\left(\alpha, \lambda\right)$

$$f\left(x\right) = \frac{x^{\alpha-1}\lambda^{\alpha}}{\Gamma\left(\alpha\right)}e^{-\lambda x}, \qquad x > 0, \quad \alpha > 0, \quad \lambda > 0$$

where

$$\Gamma\left(\alpha\right) = \int_0^{\infty} t^{\alpha-1}e^{-t}dt$$

is the Gamma function.

It can be shown that

$$\Gamma(\alpha+1) = \alpha\Gamma(\alpha), \quad \text{for all} \quad \alpha > 0$$

$$\Gamma(n+1) = n! \quad \text{for all integer } n \geq 1$$

$$\Gamma(1/2) = \sqrt{\pi}$$

Density Integrates to One

$$I = \int_0^\infty \frac{x^{\alpha-1}\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-\lambda x} dx$$

Change of variable:

$$y = \lambda x, \quad x = \frac{y}{\lambda}, \quad dx = \frac{dy}{\lambda}$$

$$I = \frac{\lambda^\alpha}{\lambda\Gamma(\alpha)} \int_0^\infty \left(\frac{y}{\lambda}\right)^{\alpha-1} e^{-y} dy$$

Therefore,

$$
\begin{aligned}
I &= \frac{\lambda^{\alpha}}{\lambda^{\alpha}\Gamma(\alpha)} \int_0^{\infty} y^{\alpha-1} e^{-y} dy \\
&= \frac{\lambda^{\alpha}\Gamma(\alpha)}{\lambda^{\alpha}\Gamma(\alpha)} = 1
\end{aligned}
$$

Mean

$$
\begin{aligned}
E(X) &= \int_0^\infty x \frac{x^{\alpha-1}\lambda^\alpha}{\Gamma(\alpha)} e^{-\lambda x} dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^\infty x^\alpha e^{-\lambda x} dx \\
&= \frac{\lambda^\alpha}{\Gamma(\alpha)} \frac{\Gamma(\alpha+1)}{\lambda^{\alpha+1}} \int_0^\infty \frac{x^\alpha e^{-\lambda x} \lambda^{\alpha+1}}{\Gamma(\alpha+1)} dx \\
&= \frac{1}{\lambda} \frac{\alpha \Gamma(\alpha)}{\Gamma(\alpha)} = \frac{\alpha}{\lambda}.
\end{aligned}
$$

Inverse Scale Parameter

Remark: if

$$Y = tX, \qquad t > 0$$

then

$$Y \sim Gamma\left(\alpha, \lambda/t\right).$$

Therefore λ is an "inverse scale" parameter (or $1/\lambda$ is a scale parameter).

Variance

It can be shown that (students should verify this)

$$Var\left(X\right) = \frac{\alpha}{\lambda^2}$$

Easier approach to show this: Use "Moment generating Function" (introduced below)

Some Particular Cases

- **Exponential distribution** $X \sim Exp(\lambda)$:

$$\alpha = 1$$

- **Chi-Square distribution with** n **degrees of freedom** $X \sim \chi_{(n)}$:

$$\alpha = n/2, \quad \lambda = 1/2$$

- **Beta distribution:** if $X \sim Gamma\left(\alpha, \lambda\right)$ and $Y \sim Gamma\left(\beta, \lambda\right)$ are independent then

$$R = X/\left(X+Y\right) \sim Beta\left(\alpha, \beta\right)$$

$$f\left(r\right) = \frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)} r^{\alpha-1}\left(1-r\right)^{\beta-1}, \qquad 0 < r < 1.$$

Important Property: we will show later (using moment generating functions) that if

$$X_i \sim Gamma\left(\alpha_i, \lambda\right), \quad i = 1, ..., n$$

are independent then

$$S = \sum_{i=1}^{n} X_i \sim Gamma\left(\sum_{i=1}^{n} \alpha_i, \lambda\right).$$

RANDOM VECTORS

Definition: Let $(\Omega, \mathcal{F}, P)$ be a probability space.

$$\mathbf{X} : \Omega \rightarrow \mathcal{R}^m$$

such that

$$\mathbf{X}^{-1}(B) \in \mathcal{F}, \text{ for all } B \in \mathcal{B}(R^m) \quad \text{(measurable function)}$$

Induced probability:

$$P_{\mathbf{X}}(B) = P\left(\mathbf{X}^{-1}(B)\right), \quad B \in \mathcal{B}(R^m).$$

Joint distribution function:

$$F(\mathbf{x}) \quad = \quad P(\mathbf{X} \leq \mathbf{x}) = P(X_1 \leq x_1, ..., X_m \leq x_m)$$

Again, the joint distribution function completely determines $P_{\mathbf{X}}(B)$

Discrete Random Vectors

The range of $\mathbf{X}$ is finite or conutable:

$$\mathbf{X}(\Omega) = \{\mathbf{x}_i\}_{i \in I}$$

$$I = \{1, 2, \ldots n\} \quad \text{or} \quad I = \mathcal{N}^+ \text{ (positive integer numbers)}$$

Joint Probability Mass function

$$
\begin{aligned}
f(\mathbf{x}_i) &= P(\mathbf{X} = \mathbf{x}_i) \\
&= P(X_1 = x_1, X_2 = x_2, ..., X_m = x_m)
\end{aligned}
$$

Then

$$
F(\mathbf{x}) = \sum_{\mathbf{x}_i \leq \mathbf{x}} \cdots \sum f(\mathbf{x}_i)
$$

Definition of joint density function: $f(\mathbf{x})$ is a function

$$f \quad : \quad R^p \longrightarrow R$$

satisfying

(1) $f(\mathbf{x}) \geq 0$, for all $\mathbf{x} \in R^p$, and

(2) $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\mathbf{x})\, dx_1 \cdots dx_p = 1.$

In this case

$$F(\mathbf{x}) = P(X_1 \leq x_1, ..., X_p \leq x_p) = \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(\mathbf{t})\, dt_1 \cdots dt_p$$

By the Fundamental Theorem of Calculus

$$f(\mathbf{x}) = \frac{\partial^p}{\partial x_1 \partial x_2 \cdots \partial x_p} F(\mathbf{x})$$

Let $\mathbf{X}$ be a continuous random vector with joint density $f_{\mathbf{X}}(\mathbf{x})$ and

$$\begin{aligned} \mathbf{y} &= \mathbf{g}(\mathbf{x}) \\ \mathbf{x} &= \mathbf{h}(\mathbf{y}) \quad \text{1-1 transformation} \end{aligned}$$

Let

$$J(\mathbf{y}) = \left| \det \left(\frac{\partial h_i(\mathbf{y})}{\partial y_j} \right) \right| \qquad \text{(Jacobian)}$$

Then

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}(\mathbf{h}(\mathbf{y})) \left| \det \left(\frac{\partial h_i(\mathbf{y})}{\partial y_j} \right) \right| = f_{\mathbf{X}}(\mathbf{h}(\mathbf{y})) J(\mathbf{y})$$

Example of Change of Variables

Suppose that

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

has uniform density on the unit circle. That is

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\pi}, \quad x_1^2 + x_2^2 \leq 1$$

Let

$$\begin{aligned} x_1 &= h_1(r, \theta) = r\cos(\theta) \\ x_2 &= h_2(r, \theta) = r\sin(\theta) \end{aligned}$$

where $0 \leq \theta < 2\pi$ and $0 \leq r \leq 1$.

Then,

$$J(r,\theta) = \left|\det\begin{pmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{pmatrix}\right| = \left|r\left(\cos^2(\theta)+\sin^2(\theta)\right)\right| = r$$

Therefore,

$$f(r,\theta) = f_{\mathbf{X}}(\mathbf{h}(r,\theta))\,J(r,\theta) = \frac{1}{\pi}r, \quad 0 \le \theta < 2\pi \text{ and } 0 \le r \le 1$$

Marginal Density

$$f\left(\mathbf{x}^{(1)}\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\right) \mathbf{dx}^{(2)} \qquad \text{(continuous case)}$$

$$f\left(\mathbf{x}^{(1)}\right) = \sum_{\mathbf{x}^{(2)}} f\left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\right) \qquad \text{(discrete case)}$$

Conditional Density

$$f\left(\mathbf{x}^{(1)}|\mathbf{x}^{(2)}\right)=\frac{f\left(\mathbf{x}^{(1)},\mathbf{x}^{(2)}\right)}{f\left(\mathbf{x}^{(2)}\right)},\qquad \text{(conditional)}$$

Example (continued)

$$f(\theta) = \int_0^1 \frac{1}{\pi} r dr = \frac{1}{\pi} \left. \frac{r^2}{2} \right|_0^1 = \frac{1}{2\pi}, \quad 0 \leq \theta < 2\pi$$

$$f(r|\theta) = \frac{\frac{1}{\pi} r}{\frac{1}{2\pi}} = 2r, \quad 0 \leq r \leq 1.$$

Independence

We say that $X_1, X_2, ..., X_p$ are independent if

$$f\left(x_1, x_2, ..., x_p\right) = \prod_{i=1}^{p} f_{X_i}\left(x_i\right)$$

where $f_{X_i}\left(x_i\right)$ is the marginal density of X_i $(i = 1, 2, ..., p)$.

Note: $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(2)}$ are independent if

$$f\left(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}\right) = f_{\mathbf{X}^{(1)}}\left(\mathbf{x}^{(1)}\right) f_{\mathbf{X}^{(2)}}\left(\mathbf{x}^{(2)}\right)$$

It is immediate that in this case

$$f\left(\mathbf{x}^{(1)} \mid \mathbf{x}^{(2)}\right) = f_{\mathbf{X}^{(1)}}\left(\mathbf{x}^{(1)}\right) \quad \text{and} \quad f\left(\mathbf{x}^{(2)} \mid \mathbf{x}^{(1)}\right) = f_{\mathbf{X}^{(2)}}\left(\mathbf{x}^{(2)}\right)$$

Change of Variables (Many-to-one transformations)

Let $\mathbf{X}$ be a continuous random vector with joint density $f_{\mathbf{X}}(\mathbf{x})$. Let

$$\mathcal{X} = \{\mathbf{x} : \; f_{\mathbf{X}}(\mathbf{x}) > 0\} \qquad \text{(support space for } f_{\mathbf{X}}\text{)}$$

Consider

$$\begin{aligned} \mathbf{g} \quad &: \quad \mathcal{X} \to \mathcal{R} \subset R^p \\ \mathbf{y} \quad &= \quad \mathbf{g}(\mathbf{x}) \qquad \text{many-to-one, onto } \mathcal{R} \end{aligned}$$

Assume that $\mathcal{X}$ can be decomposed into sets

$$\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_m$$

such that

$$\mathbf{g} : \mathcal{X}_i \to \mathcal{R} \quad \text{one-to-one,} \;\; \text{onto } \mathcal{R} \quad (i = 1, ..., m)$$

Let

$$\mathbf{h}_i : \mathcal{R} \to \mathcal{X}_i \quad \text{inverses} \quad i = 1, ..., m$$

That is

$$\mathbf{h}_i \left(\mathbf{g} \left(\mathbf{x} \right) \right) = \mathbf{x}, \quad \text{for all} \;\; \mathbf{x} \in \mathcal{X}_i$$

$$\text{Set} \quad J_i(\mathbf{y}) = \left| \det \left(\frac{\partial \mathbf{h}_i(\mathbf{y})}{\partial \mathbf{y}} \right) \right| \qquad \text{(Jacobians)}$$

Then, for all $\mathbf{y} \in \mathcal{R}$,

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i=1}^{m} f_{\mathbf{X}}(\mathbf{h}_i(\mathbf{y})) \left| \det \left(\frac{\partial \mathbf{h}_i(\mathbf{y})}{\partial \mathbf{y}} \right) \right| = \sum_{i=1}^{m} f_{\mathbf{X}}(\mathbf{h}_i(\mathbf{y})) J_i(\mathbf{y}),$$

provided that

- all the partial derivatives in $J_i(\mathbf{y})$ are continuous
- and all the determinants are non-zero

Example 1:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

$$y = g(x) = x^2$$

$$\mathcal{R} = [\ 0, \infty), \quad \mathcal{X}_1 = (-\infty, 0) \quad \text{and} \quad \mathcal{X}_2 = (0, \infty)$$

(make a picture to better understand this)

$$h_1 : (0,\infty) \to (-\infty, 0)$$

$$h_1(y) = -\sqrt{y}, \qquad J_1(y) = \frac{y^{-1/2}}{2}$$

$$h_2 : (0,\infty) \to (0, \infty)$$

$$h_2(y) = \sqrt{y} \qquad J_2(y) = \frac{y^{-1/2}}{2}$$

Therefore,

$$\begin{aligned}
f_Y(y) &= f_X(h_1(y))\,J_1(y) + f_X(h_2(y))\,J_2(y) \\
&= f_X(-\sqrt{y})\,\frac{y^{-1/2}}{2} + f_X(\sqrt{y})\,\frac{y^{-1/2}}{2} = \frac{y^{1/2-1}}{\sqrt{2\pi}}e^{-y/2}
\end{aligned}$$

Chi-square with one d.f.

Exercise 1

Exercise 1: Let

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

and

$$y = g(x) = \begin{cases} x^2 & x < 0 \\ x & x > 0 \end{cases}$$

Find $f_Y(y)$.

Example 2

Example 2:

$$f_X(x_1, x_2, x_3) = \frac{3}{4\pi}, \quad x_1^2 + x_2^2 + x_3^2 \le 1$$

Find the density for

$$\mathbf{g}(\mathbf{x}) = \left(x_{(1)}, x_{(2)}, x_{(3)}\right) = (y_1, y_2, y_3).$$

That is, $\mathbf{y} = (y_1, y_2, y_3)$ is equal to the sorted vector $\mathbf{x}$.

Solution

$$
\begin{aligned}
\mathcal{R} &= \left\{ (y_1, y_2, y_3)\,, \quad y_1^2 + y_2^2 + y_3^2 \leq 1, \quad y_1 < y_2 < y_3 \right\} \\
\mathcal{X}_1 &= \left\{ (x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \leq 1, \quad x_1 < x_2 < x_3 \right\} \\
\mathcal{X}_2 &= \left\{ (x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \leq 1, \quad x_1 < x_3 < x_2 \right\} \\
\mathcal{X}_3 &= \left\{ (x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \leq 1, \quad x_2 < x_1 < x_3 \right\} \\
\mathcal{X}_4 &= \left\{ (x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \leq 1, \quad x_2 < x_3 < x_1 \right\} \\
\mathcal{X}_5 &= \left\{ (x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \leq 1, \quad x_3 < x_1 < x_2 \right\} \\
\mathcal{X}_6 &= \left\{ (x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \leq 1, \quad x_3 < x_2 < x_1 \right\}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{h}_1 &: \mathcal{R} \to \mathcal{X}_1, & \mathbf{h}_1(y_1, y_2, y_3) &= (y_1, y_2, y_3), & J_1(\mathbf{y}) &= 1 \\
\mathbf{h}_2 &: \mathcal{R} \to \mathcal{X}_2, & \mathbf{h}_1(y_1, y_2, y_3) &= (y_1, y_3, y_2), & J_2(\mathbf{y}) &= 1 \\
\mathbf{h}_3 &: \mathcal{R} \to \mathcal{X}_3, & \mathbf{h}_1(y_1, y_2, y_3) &= (y_2, y_1, y_3), & J_3(\mathbf{y}) &= 1 \\
\mathbf{h}_4 &: \mathcal{R} \to \mathcal{X}_4, & \mathbf{h}_1(y_1, y_2, y_3) &= (y_2, y_3, y_1), & J_4(\mathbf{y}) &= 1 \\
\mathbf{h}_5 &: \mathcal{R} \to \mathcal{X}_5, & \mathbf{h}_1(y_1, y_2, y_3) &= (y_3, y_1, y_2), & J_5(\mathbf{y}) &= 1 \\
\mathbf{h}_6 &: \mathcal{R} \to \mathcal{X}_6, & \mathbf{h}_1(y_1, y_2, y_3) &= (y_3, y_2, y_1), & J_6(\mathbf{y}) &= 1
\end{aligned}
$$

Therefore

$$
\begin{aligned}
f_{\mathbf{Y}}(\mathbf{y}) &= f(y_1, y_2, y_3) + f(y_1, y_3, y_2) + \cdots + f(y_3, y_2, y_1) \\
&= \frac{6 \times 3}{4\pi} = \frac{4.5}{\pi}, \quad \text{for} \;\; y_1^2 + y_2^2 + y_3^2 \leq 1, \;\; y_1 < y_2 < y_3
\end{aligned}
$$

Exercise 2

Suppose that $(X_1, X_2, ..., X_n)$ are i.i.d. with common uniform distribution on $(0, 1)$.

(a) Find the joint density for the order statistics

$$\mathbf{Y} = \left(X_{(1)}, X_{(2)}, ..., X_{(n)}\right)$$

(b) Find the marginal density for $X_{(i)}$, $i = 1, ..., n$.

(c) Find the density for the range $X_{(n)} - X_{(1)}$.

PROPERTIES OF THE EXPECTED VALUE

Linearity

- $Y = a + bX \Rightarrow E(Y) = a + bE(X)$ [easy to prove]

- $E(aX + bY) = aE(X) + bE(Y)$ [easy to prove]

Follow directly from linearity of the integral (or sum)

Monotonicity

- $X \geq 0$ and $E(X) = 0 \Rightarrow X = 0$, outside a set of probability zero.

- $E(|X|) < \infty \Rightarrow |X| < \infty$, outside a set of probability zero.

  [These are intuitively clear but non-trivial results]

Monotonicity

- $X \leq Y \Rightarrow \; E(X) \leq E(Y)$.

- **Monotone Convergence Theorem (MCT)**:

  $0 \leq X_n \uparrow X$ a.s. $\Rightarrow 0 \leq E(X_n) \uparrow E(X)$
  $[\lim_{n\to\infty} E(X_n) = E(\lim_{n\to\infty} X_n)]$

Dominated Convergence Theorem

- $X_n \to X$ a.s. and $|X_n| \leq Y$, a.s. with $E(|Y|) < \infty \Rightarrow$

  (1) $E(|X|) < \infty$ and so $E(X)$ is well defined and finite. In addition

  (2) $\lim_{n\to\infty} E(X_n) = E(\lim_{n\to\infty} X_n) = E(X)$

- $g(X, Y) \geq 0 \Rightarrow$

$$E(g(X, Y)) = E_Y(E_X(g(X, Y))) = E_X(E_Y(g(X, Y)))$$

- Either $E_Y(E_X(|g(X, Y)|)) < \infty$ or $E_X(E_Y(|g(X, Y)|)) < \infty \Rightarrow$

$$E(g(X, Y)) = E_Y(E_X(g(X, Y))) = E_X(E_Y(g(X, Y)))$$

Independence

- X and Y are independent, $E(|X|) < \infty$ and $E(|Y|) < \infty \Rightarrow$

$$E(XY) = E(X)E(Y)$$

Proof: By Tonelli's Theorem:

$$
\begin{aligned}
E\left(|XY|\right) &= E_Y\left(E_X\left(|XY|\right)\right) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}|x|\,|y|\,f_X\left(x\right)dx\right]f_Y\left(y\right)dy \\
&= \int_{-\infty}^{\infty}\left[|y|\,f_Y\left(y\right)\int_{-\infty}^{\infty}|x|\,f_X\left(x\right)dx\right]dy \\
&= \int_{-\infty}^{\infty}\left[|y|\,f_Y\left(y\right)E\left(|X|\right)\right]dy \\
&= E\left(|X|\right)\int_{-\infty}^{\infty}|y|\,f_Y\left(y\right)dy \\
&= E\left(|X|\right)E\left(|Y|\right) < \infty
\end{aligned}
$$

Now, using Fubini's Theorem in a similar way gives the desired result.

Since $E_Y\left(E_X\left(|XY|\right)\right) < \infty$, by Fubinis's Theorem

$$\begin{aligned}
E\left(XY\right) &= E_Y\left(E_X\left(XY\right)\right) = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} xyf_X\left(x\right)dx\right]f_Y\left(y\right)dy \\
&= \int_{-\infty}^{\infty}\left[yf_Y\left(y\right)\int_{-\infty}^{\infty} xf_X\left(x\right)dx\right]dy \\
&= \int_{-\infty}^{\infty}\left[yf_Y\left(y\right)E\left(X\right)\right]dy \\
&= E\left(X\right)\int_{-\infty}^{\infty} yf_Y\left(y\right)dy \\
&= E\left(X\right)E\left(Y\right)
\end{aligned}$$

Example 1: Random Series

Let X_n be a sequence of random variables. Set

$$S_n = \sum_{i=1}^{n} X_i \quad , \quad Y_n = \sum_{i=1}^{n} |X_i| \quad \text{and} \quad Y = \sum_{i=1}^{\infty} |X_i|$$

Then we have

Par 1. $E\left(\sum_{i=1}^{\infty} |X_i|\right) = \sum_{i=1}^{\infty} E\left(|X_i|\right)$

Part 2. If in addition $\sum_{i=1}^{\infty} E\ \left(|X_i|\right) < \infty$ then:

(a) $\sum_{i=1}^{n} X_i$ converges absolutely a.s. and

(b) $E\left(\sum_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} E\left(X_i\right)$

Proof of Part 1: By monotonicity

$$\sum_{i=1}^{n} |X_i| = Y_n \uparrow Y = \lim_{n\to\infty} \sum_{i=1}^{n} |X_i| = \sum_{i=1}^{\infty} |X_i| \, .$$

By the MCT

$$\lim_{n\to\infty} E\left(Y_n\right) = E\left(\lim_{n\to\infty} Y_n\right) = E\left(Y\right) = E\left(\sum_{i=1}^{\infty} |X_i|\right) \tag{1}$$

proving Part 1.

Remark: From (1)

$$
\begin{aligned}
E\left(\sum_{i=1}^{\infty}|X_i|\right) &= \lim_{n\to\infty} E\left(Y_n\right) \\
&= \lim_{n\to\infty} E\left(\sum_{i=1}^{n}|X_i|\right) = \lim_{n\to\infty}\sum_{i=1}^{n} E\left(|X_i|\right) \\
&= \sum_{i=1}^{\infty} E\left(|X_i|\right),
\end{aligned}
$$

Proof of Part 2 (a): From Part 1, $E\left(\sum_{i=1}^{\infty}|X_i|\right) < \infty$ and so

$$\sum_{i=1}^{\infty}|X_i| < \infty \qquad \text{a.s.}$$

$$\Rightarrow \sum_{i=1}^{n} X_i \quad \text{converges absolutely a.s.} \tag{2}$$

proving Par 2 (a)

Proof of Part 2 (b): By (2)

$$\sum_{i=1}^{n} X_i \to \sum_{i=1}^{\infty} X_i .$$

and for all n,

$$\left| \sum_{i=1}^{n} X_i \right| \leq \sum_{i=1}^{n} |X_i| < Y, \quad \text{a.s.} \quad \text{with} \quad E\left(Y\right) < \infty.$$

By the DCT

and

$$E\left(\left|\sum_{i=1}^{\infty} X_i\right|\right) < \infty \quad \text{and} \quad \lim_{n\to\infty} E\left(\sum_{i=1}^{n} X_i\right) = E\left(\sum_{i=1}^{\infty} X_i\right) \tag{3}$$

Notice that

$$\lim_{n\to\infty} E\left(\sum_{i=1}^{n} X_i\right) = \lim_{n\to\infty} \sum_{i=1}^{n} E\left(X_i\right) = \sum_{i=1}^{\infty} E\left(X_i\right) \tag{4}$$

From (3) and (4)

$$E\left(\sum_{i=1}^{\infty} X_i\right) = \sum_{i=1}^{\infty} E\left(X_i\right), \quad \text{proving Part 2 (b)}$$

Example 2: Differentiating Expected Values

Let $X(t, w)$ be a random function (random process).

- For fixed $w \in \Omega$, $X(t, w)$ is a (deterministic) function of $t \in (a, b)$.
- For fixed $t \in (a, b)$, $X(t, w)$ is a random variable on $(\Omega, \mathcal{F}, P)$

Suppose that:

(1) There exists $t_0 \in (a, b)$ such that $E(|X(t_0)|) < \infty$

(2) $X'(t, w) = \frac{d}{dt}X(t, w)$ exist for all $t \in (a, b)$ a.s. [P] and

(3) There exists a random variable $W(w)$, with $E(|W|) < \infty$, such that

$$\left|X'(t, w)\right| \leq W(w), \quad \text{for all } t \in (a, b) \quad \text{a.s.} \quad [P]$$

Then:

(a) $E\left(\left|X\left(t\right)\right|\right) < \infty, \qquad$ for all $\ t \in \left(a, b\right)$

(b) $E\left(\left|X'\left(t\right)\right|\right) < \infty, \quad$ tor all $t \in \left(a, b\right)$

(c) $\frac{d}{dt}E\left(X\left(t\right)\right) = E\left(\frac{d}{dt}X\left(t\right)\right).$

Proof:
(a) Left as an exercise.

(b) Let $t \in (a, b)$ be fixed.

Consider a sequence $h_n \to 0$ (with $h_n \neq 0$), such that $t + h_n \in (a, b)$ for all n.

By the mean value theorem

$$X(t + h_n) - X(t) = h_n X'(s_n)$$

where $s_n = \alpha_n t + (1 - \alpha_n)(t + h_n)$, with $0 \leq \alpha_n \leq 1$. Set

$$X'(s_n) = \frac{X(t + h_n) - X(t)}{h_n} = Y_n(t).$$

$$X'(s_n) = \frac{X(t+h_n) - X(t)}{h_n} = Y_n(t).$$

By assumption, $Y_n(t) \to X'(t)$, a.s. $[P]$ and $|Y_n(t)| \leq W$.

By the DCT

$$E\left(\left|X'(t)\right|\right) < \infty$$

and

$$\lim_{n\to\infty} E(Y_n(t)) = E\left(\lim_{n\to\infty} Y_n(t)\right) = E\left(X'(t)\right)$$

(c) We have just shown that

$$\lim_{n \to \infty} E\left(Y_n\left(t\right)\right) = E\left(\frac{d}{dt} X\left(t\right)\right) \tag{5}$$

Moreover, by definition of derivative,

$$\lim_{n \to \infty} E\left(Y_n\left(t\right)\right) = \lim_{n \to \infty} \frac{E\left[X\left(t + h_n\right)\right] - E\left[X\left(t\right)\right]}{h_n} = \frac{d}{dt} E\left(X\left(t\right)\right) \tag{6}$$

The result follows now from (5) and (6).

MULTIVARIATE MEAN AND COVARIANCE MATRIX

Multivariate Mean and Covariance Matrix

$$\mu = E\left(\mathbf{X}\right)$$

$$\Sigma = E\left\{\left(\mathbf{X} - \mu\right)\left(\mathbf{X} - \mu\right)'\right\}$$

$$R = D^{-1/2}\Sigma D^{-1/2}, \qquad D = diag\left(\Sigma\right)$$

It is easy to show that

$$\Sigma = E\left\{\mathbf{X}\mathbf{X}'\right\} - \boldsymbol{\mu}\boldsymbol{\mu}'$$

Moreover, if

$$\mathbf{Y} = A\mathbf{X} + \mathbf{b}$$

then

$$E\left(\mathbf{Y}\right) = AE\left(\mathbf{X}\right) + \mathbf{b} = A\boldsymbol{\mu} + \mathbf{b}$$

$$Cov\left(\mathbf{Y}\right) = ACov\left(\mathbf{X}\right)A' = A\Sigma A'$$

- If

$$Y = \mathbf{X}'\mathbf{a}$$

then

$$E(Y) = \mathbf{a}'\boldsymbol{\mu} \quad , \quad Var(Y) = \mathbf{a}\Sigma\mathbf{a}'$$

- Suppose that A is symmetric ($A' = A$) and

$$Y = \mathbf{X}'A\mathbf{X}$$

then

$$E(Y) = \boldsymbol{\mu}'A\boldsymbol{\mu} + tr(A\Sigma)$$

Proof

$$
\begin{aligned}
E\left(Y\right) &= E\left(\mathbf{X}'A\mathbf{X}\right) \\
&= E\left[tr\left(\mathbf{X}'A\mathbf{X}\right)\right] = E\left[tr\left(\mathbf{X}\mathbf{X}'A\right)\right] \\
&= trE\left[\left(\mathbf{X}\mathbf{X}'A\right)\right] = tr\left[E\left(\mathbf{X}\mathbf{X}'\right)A\right] \\
&= tr\left[\left(\Sigma + \mu\mu'\right)A\right] = tr\left(\Sigma A\right) + tr\left(\mu\mu'A\right) \\
&= tr\left(\Sigma A\right) + tr\left(\mu'A\mu\right) = tr\left(\Sigma A\right) + \mu'A\mu
\end{aligned}
$$

SOME INEQUALITIES

Chevichev's and Markov's Inequalities

$$P\left(|X-\mu|>\epsilon\right)\leq\frac{\sigma^2}{\epsilon^2}$$

$$P\left(|X-\mu|>\epsilon\right)\leq\frac{E\left|X-\mu\right|}{\epsilon}$$

The proof of these two inequalities are left as exercise.

Jensen's Inequality

If $g(x)$ is a convex function then

$$E(g(X)) \geq g(E(X))$$

Proof of Jensen's Inequality

Let $l(x) = a + bx$ such that

$$g(x) \geq a + bx, \quad \text{for all } x$$

and

$$g(\mu) = a + b\mu$$

This linear function exists because $g(x)$ is convex ($l(x)$ is a support line). Now,

$$E(g(X)) \geq E(a + bX) = a + b\mu = g(\mu).$$

Cauchy-Schwarz Inequality

(1)

$$Cov^2(X, Y) \leq Var(X)\, Var(Y)$$

(2) Equality holds if and only if $Y = t_0 + t_1 X$, for some constants t_0 and t_1.

Proof. Let

$$g(t) \;=\; Var(Y - tX) = Var(Y) - 2Cov(X, Y)\,t + Var(X)\,t^2$$

Notice that:

- $g(t)$ is a convex, second degree polynomial in t.
- $g(t)$ has a unique root or no root at all (it cannot be negative).

Recall the formula for the roots of a second degree polynomial:

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-b \pm \Delta}{2a}.$$

Notice that the discriminant is

$$\Delta^2 = b^2 - 4ac \leq 0.$$

In our case

$$a = Var\left(Y\right), \quad b = -2Cov\left(X, Y\right) \quad \text{and} \quad c = Var\left(X\right)$$

Therefore,

$$\Delta^2 = 4Cov^2\left(X, Y\right) - 4Var\left(Y\right)Var\left(X\right) \leq 0,$$

proving Part 1.

Proof of Part 2:

- If $g(t) = 0$ has a solution, t_0,say then $Var(Y - t_0 X) = 0$
- This means that for some t_0 that is,

$$Y - t_0 X = t_1 \quad \text{a.s.}$$

- In other words, equality holds if and only if

$$Y = t_0 X + t_1, \qquad \text{a.s.} \qquad (Y \text{ is a linear function of } X)$$

The Pearson Correlation Coefficient

$$\rho = \frac{Cov(X, Y)}{\sqrt{Var(X)\, Var(Y)}}$$

By Cauchy-Schwarz inequality

$$|\rho| \leq 1,$$

with equality if and only if

$$Y = aX + b, \quad \text{a.s.}$$

MOMENTS AND MOMENT GENERATING FUNCTION

Definitions

$$\mu_k = E\left(X^k\right) \qquad \text{(moments)}$$

$$M_X(t) = E(\exp(tX)) \qquad \text{(mgf)}$$

Result 1: if $M_X(t)$ exists in an open interval around zero, then $M_X(t)$ is infinitely differentiable at zero and

$$\frac{d^k}{dt^k} M_X(t)\bigg|_{t=0} = \frac{d^k}{dt^k} M_X(0) = \mu_k$$

Proof: Use the result of Example 2 and the fact that, for all k,

$$|X| \exp(tX) \leq C_k + \exp(2tX) + \exp(-2tX),$$

which has finite mean for t small enough.

Result 2: if $M_X(t) = M_Y(t)$ for all t in an open interval around zero, then X and Y have the same distribution.

Result 3: if $X_1, X_2, ..., X_n$ are independent random variables with m.g.f. $M_{X_i}(t)$, then

$$M_{\sum X_i}(t) = \prod M_{X_i}(t)$$

Proof:

$$M_{\sum X_i}(t) = E\left(\exp\left\{t \sum X_i\right\}\right) = E\left(\prod \exp\left\{tX_i\right\}\right)$$

$$= \prod E\left(\exp\left\{tX_i\right\}\right) = \prod M_{X_i}(t)$$

Result 4: If $Y = a + bX$, then

$$M_Y(t) = e^{ta} M_X(bt)$$

Proof:

$$M_Y(t) = E\left(e^{tY}\right) = E\left(e^{t(a+bX)}\right)$$

$$= E\left(e^{ta} e^{tbX}\right) = e^{ta} E\left(e^{tbX}\right)$$

$$= e^{ta} M_X(bt)$$

MOMENT GENERATING FUNCTION FOR SOME RANDOM VARIABLES

Binomial

$$X \sim B(n, p)$$

$$X = X_1 + X_2 + \cdots + X_n, \qquad \text{i.i.d. } B(1, p)$$

$$M_{X_1}(t) = 1 + p\left(e^t - 1\right)$$

$$M_{X_1+X_2+\cdots+X_n}(t) = \left[M_{X_1}(t)\right]^n = \left(1 + p\left(e^t - 1\right)\right)^n$$

Poisson

$$X \sim \mathcal{P}(\lambda)$$

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx} e^{-\lambda} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!}$$

$$= \exp(-\lambda) \exp\left\{\lambda e^t\right\} = \exp\left\{\lambda\left(e^t - 1\right)\right\}$$

Geometric

$$X \sim Geometric\left(p\right)$$

$$M_X\left(t\right) = \sum_{x=1}^{\infty} e^{xt} p\left(1-p\right)^x = p \sum_{x=1}^{\infty} \left[e^t\left(1-p\right)\right]^x,$$

$$= \frac{p}{1-e^t\left(1-p\right)}, \quad \text{provided} \quad t < -\log\left(1-p\right)$$

Standard Normal

$$Z \sim N(0,1)$$

$$M_Z(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tz} e^{-\frac{1}{2}z^2} dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(z^2 - 2tz\right)} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(z^2 - 2tz + t^2\right) + \frac{1}{2}t^2} dz = \exp\left(\frac{1}{2}t^2\right)$$

General Normal

In the general case

$$
\begin{aligned}
X &\sim N\left(\mu, \sigma^2\right) \\
X &= \mu + \sigma Z
\end{aligned}
$$

$$
M_{\mu+\sigma Z}\left(t\right) = e^{t\mu} M_Z\left(\sigma t\right) = \exp\left(t\mu + \frac{1}{2} t^2 \sigma^2\right)
$$

Gamma

$$X \sim Gamma\left(\alpha, \lambda\right)$$

Let $t < \lambda$,

$$M_X\left(t\right) = \int_0^\infty e^{tx} \frac{x^{\alpha-1}\lambda^\alpha}{\Gamma\left(\alpha\right)} e^{-\lambda x} dx = \int_0^\infty \frac{x^{\alpha-1}\lambda^\alpha}{\Gamma\left(\alpha\right)} e^{-(\lambda-t)x} dx$$

$$= \left(\lambda - t\right)^{-\alpha} \lambda^\alpha \int_0^\infty \frac{x^{\alpha-1}}{\Gamma\left(\alpha\right)} \left(\lambda - t\right)^\alpha e^{-(\lambda-t)x} dx$$

$$= \left(\frac{\lambda - t}{\lambda}\right)^{-\alpha} = \left(1 - \left(t/\lambda\right)\right)^{-\alpha}, \quad \text{for} \quad t < \lambda$$

Mean and Variance for Gamma

$$M_X(t) \quad = \quad (1 - (t/\lambda))^{-\alpha}$$

$$E(X) = M'(t)\Big|_{t=0} = (\alpha/\lambda)(1-(t/\lambda))^{-(\alpha+1)}\Big|_{t=0} = \frac{\alpha}{\lambda}$$

$$E\left(X^2\right) = M''(t)\Big|_{t=0} = \frac{(\alpha+1)\,\alpha}{\lambda^2}(1-(t/\lambda))^{-\alpha-1}\Big|_{t=0} = \frac{(\alpha+1)\,\alpha}{\lambda^2}$$

Therefore,

$$
\begin{aligned}
E(X) &= \frac{\alpha}{\lambda} \\
Var(X) &= \frac{\alpha(\alpha+1)}{\lambda^2} - \frac{\alpha^2}{\lambda^2} = \frac{\alpha}{\lambda^2}
\end{aligned}
$$

Particular case: $X \sim \chi^2_{(n)} \quad [\alpha = n/2 \text{ and } \lambda = 1/2]$

$$
E(X) = \frac{n/2}{1/2} = n, \quad Var(X) = \frac{n/2}{1/4} = 2n
$$

MOMENT GENERATING FUNCTION FOR RANDOM VECTORS

Definition

$$
\begin{aligned}
M_{\mathbf{X}}(\mathbf{t}) &= E\left(\exp\left(\mathbf{t}'\mathbf{X}\right)\right) \\
&= E\left(\exp\left(t_1 X_1 + \cdots + t_p X_p\right)\right)
\end{aligned}
$$

Result 1

If $M_{\mathbf{X}}(\mathbf{t})$ exists in an **open interval that includes 0**, then $M_{\mathbf{X}}(\mathbf{t})$ is **infinitely differentiable** at zero and

$$\left.\frac{\partial}{\partial t_1^{k_1} \ldots \partial t_m^{k_m}} M_{\mathbf{X}}(\mathbf{t})\right|_{\mathbf{t}=\mathbf{0}} = \frac{\partial}{\partial t_1^{k_1} \ldots \partial t_m^{k_m}} M_{\mathbf{X}}(0) = E\left(\prod_{i=1}^{m} X_i^{k_i}\right)$$

For instance

$$E(X_1 X_2) = \left.\frac{\partial}{\partial t_1 \partial t_2} M_{\mathbf{X}}(\mathbf{t})\right|_{\mathbf{t}=\mathbf{0}}$$

and

$$E\left(X_1 X_2^2\right) = \left.\frac{\partial}{\partial t_1 \partial t_2^2} M_{\mathbf{X}}(\mathbf{t})\right|_{\mathbf{t}=\mathbf{0}}$$

Result 2

If for some $\epsilon > 0$,

$$M_{\mathbf{X}}(\mathbf{t}) = M_{\mathbf{Y}}(\mathbf{t}) \quad \text{for all } \mathbf{t} \in B(\mathbf{0}, \epsilon)$$

then

$$F_{\mathbf{X}}(\mathbf{v}) = F_{\mathbf{Y}}(\mathbf{v}) \quad \text{for all } \mathbf{v}$$

Result 3

If $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_n$ are independent random vectors with m.g.f. $M_{\mathbf{X}_i}(\mathbf{t})$, then

$$M_{\sum_{i=1}^n \mathbf{X}_i}(\mathbf{t}) = \prod_{i=1}^n M_{\mathbf{X}_i}(\mathbf{t})$$

Proof:

$$M_{\sum \mathbf{X}_i}(\mathbf{t}) = E\left(\exp\left\{\mathbf{t}' \sum \mathbf{X}_i\right\}\right) = E\left(\prod \exp\left\{\mathbf{t}'\mathbf{X}_i\right\}\right)$$

$$= \prod E\left(\exp\left\{\mathbf{t}'\mathbf{X}_i\right\}\right) = \prod M_{\mathbf{X}_i}(\mathbf{t})$$

Result 4

If

$$\mathbf{Y} = \mathbf{a} + B\mathbf{X}$$

for some $q \times p$ matrix B, then

$$M_{\mathbf{Y}}(\mathbf{t}) = \exp(\mathbf{t}'\mathbf{a})\, M_X(B'\mathbf{t})$$

Proof:

$$M_{\mathbf{Y}}(\mathbf{t}) = E\left(e^{\mathbf{t}'\mathbf{Y}}\right) = E\left(e^{\mathbf{t}'(\mathbf{a}+B\mathbf{X})}\right)$$

$$= E\left(e^{\mathbf{t}'\mathbf{a}} e^{\mathbf{t}'B\mathbf{X}}\right) = e^{\mathbf{t}'\mathbf{a}} E\left(e^{(B'\mathbf{t})'\mathbf{X}}\right)$$

$$= \exp(\mathbf{t}'\mathbf{a})\, M_{\mathbf{X}}(B'\mathbf{t})$$

MGF For The Standard Multivariate Normal

$$\mathbf{Z} \sim N(\mathbf{0}, I)$$

$$\begin{aligned}
M_{\mathbf{Z}}(\mathbf{t}) &= E\left(\exp\left\{\mathbf{t}'\mathbf{Z}\right\}\right) = E\left(\prod_{i=1}^{p} \exp\left(t_i Z_i\right)\right) \\
&= \prod_{i=1}^{p} E\left(\exp\left(t_i Z_i\right)\right) = \prod_{i=1}^{p} \exp\left(\frac{t_i^2}{2}\right) \\
&= \exp\left(\frac{1}{2}\sum_{i=1}^{p} t_i^2\right) = \exp\left(\frac{1}{2}\mathbf{t}'\mathbf{t}\right)
\end{aligned}$$

MGF For The Multivariate Normal

$$\mathbf{X} \sim N(\mu, \Sigma) \Rightarrow \mathbf{X} = \mu + \Sigma^{1/2}\mathbf{Z}$$

$$\begin{aligned} M_{\mathbf{X}}(\mathbf{t}) &= M_{\mu+\Sigma^{1/2}\mathbf{Z}}(\mathbf{t}) = \exp(\mathbf{t}'\mu)\, M_{\mathbf{Z}}\left(\Sigma^{1/2}\mathbf{t}\right) \\ &= \exp(\mathbf{t}'\mu) \exp\left\{\frac{1}{2}\left(\Sigma^{1/2}\mathbf{t}\right)' \Sigma^{1/2}\mathbf{t}\right\} \\ &= \exp\left\{\mathbf{t}'\mu + \frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right\} \end{aligned}$$