

Lecture 34: Nov 29

Last time

- Multiple Random Variables (Chapter 4)

Today

- Course Evaluations
- Conditional Distributions
- Independence

Continuous Bivariate RVs The random variables X and Y are said to be *jointly continuous* if there exists a function $f_{X,Y}(x, y)$, such that for any Borel set B of 2-tuples in \mathbb{R}^2 ,

$$\Pr\{(X, Y) \in B\} = \int \int_{(x,y) \in B} f_{X,Y}(x, y) dx dy.$$

The function $f_{X,Y}(x, y)$ is called the *joint probability density function* for X and Y . It follows in this case that

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) dt ds,$$
$$f_{X,Y}(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

Properties of the bivariate pdf

- $f_{X,Y}(x, y) \geq 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$
- $f_{X,Y}(x, y)$ is not a probability, but can be thought of as a relative probability of (X, Y) falling into a small rectangle located at (x, y) :

$$\Pr\{x < X \leq x + dx, y < Y \leq y + dy\} \approx f(x, y) dx dy$$

- The *marginal probability density functions* for X and Y can be obtained as

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$$
$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$$

Example 1

$$F_{X,Y}(x, y) = xy \quad 0 < x \leq 1, 0 < y \leq 1$$

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y} =$$

$$f_X(x) =$$

$$f_Y(y) =$$

Example 2

$$F_{X,Y}(x, y) = x - x \log \frac{x}{y} \quad 0 < x \leq y \leq 1$$

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y} =$$

$$f_X(x) =$$

$$f_Y(y) =$$

Note: Once we have $f_X(x)$ and $f_Y(y)$, we can obtain $F_X(x)$ and $F_Y(y)$ directly. Double check: $F_X(x) = F_{X,Y}(x, \infty)$.

Conditional Distributions

Conditional Distributions - Discrete Recall if A and B are two events, the probability of A conditional on B is:

$$\Pr(A|B) = \frac{\Pr(A, B)}{\Pr(B)}$$

Defining the events $A = \{Y = y\}$ and $B = \{X = x\}$, it follows that

$$\begin{aligned} \Pr\{Y = y|X = x\} &= \frac{\Pr(X = x, Y = y)}{\Pr(X = x)} \\ &= \frac{f_{X,Y}(x, y)}{f_X(x)} \\ &= f_{Y|X}(y|x) \end{aligned}$$

This is called the *conditional probability mass function* of Y given X .

Example: Discrete Back to the fair coin example. From the joint pmf of X and Y , we can derive all the conditional pmfs:

		Y			
		0	1	2	3
X	0	1/8	1/4	1/8	0
	1	0	1/8	1/4	1/8

Conditional Distribution - Continuous If $F(x, y)$ is absolutely continuous, we define the conditional density of X given Y as:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \text{ if } f_Y(y) > 0$$

Example 1

$$\begin{aligned}
F_{XY}(x, y) &= xy & 0 < x < 1, \quad 0 < y < 1 \\
f_{XY}(x, y) &= 1 & 0 < x < 1, \quad 0 < y < 1 \\
f_X(x) &= 1 & 0 < x < 1 \\
f_Y(y) &= 1 & 0 < y < 1 \\
f_{X|Y}(x|y) &= \frac{f_{XY}(x, y)}{f_Y(y)} = 1 & 0 < x < 1 \quad (0 < y < 1) \\
f_{Y|X}(y|x) &= \frac{f_{XY}(x, y)}{f_X(x)} = 1 & 0 < y < 1 \quad (0 < x < 1)
\end{aligned}$$

Note: Here we get that the conditional densities are the same as the marginals. This means X and Y are independent.

Example 2

$$\begin{aligned}
F_{XY}(x, y) &= x - x \log \frac{x}{y} & 0 < x \leq y \leq 1 \\
f_{XY}(x, y) &= 1/y & 0 < x \leq y \leq 1 \\
f_X(x) &= -\log x & 0 < x \leq 1 \\
f_Y(y) &= 1 & 0 < y \leq 1 \\
f_{X|Y}(x|y) &= \frac{f_{XY}(x, y)}{f_Y(y)} = 1/y & 0 < x \leq y \quad (0 < y \leq 1) \\
f_{Y|X}(y|x) &= \frac{f_{XY}(x, y)}{f_X(x)} = -\frac{1}{y \log x} & x \leq y \leq 1 \quad (0 < x \leq 1)
\end{aligned}$$

- Y is marginally uniform, but not conditionally uniform.
- X is conditionally uniform, but not marginally uniform.

Independent Random Variables

Independence The random variable X and Y are said to be *independent* if for any two Borel sets A and B ,

$$\Pr(X \in A, Y \in B) = \Pr(X \in A) \Pr(Y \in B)$$

All events defined in terms of X are independent of all events defined in terms of Y .

Using the Kolmogorov axioms of probability, it can be shown that X and Y are independent if and only if $\forall(x, y)$ (except possibly for sets of probability 0)

$$F_{X,Y}(x, y) = F_X(x)F_Y(y)$$

or in terms of pmfs (discrete) and pdfs (continuous)

$$f_{X,Y}(x, y) = f_X(x)f_Y(y)$$

Checking independence

- A necessary condition for independence of X and Y is that their joint pdf/pmf has positive probability on a rectangular domain.
- If the domain is rectangular, one can try to write the joint pdf/pmf as a product of functions of x and y only.

Example Two points are selected randomly on a line of length a so as to be on opposite sides of the mid-point of the line. Find the probability that the distance between them is less than $a/3$.

Solution:

Let X be the coordinate of a point selected randomly in $[0, a/2]$ and Y be the coordinate of a point selected randomly in $[a/2, a]$. Assume X and Y are independent and uniform over its interval. The joint density is

$$f_{X,Y}(x,y) = 4/a^2, \quad 0 \leq x \leq a/2, a/2 \leq y \leq a$$

Therefore, the solution is

$$\Pr(Y - X < a/3) =$$

Example: Buffon's Needle A table is ruled with lines distance 1 unit apart. A needle of length $L \leq 1$ is thrown randomly on the table. What is the probability that the needle intersects a line?

Solution:

Define two random variables:

- X : distance from low end of the needle to the nearest line above
- θ : angle from the vertical to the needle.

By “random”, we assume X and θ are independent, and

$$X \sim U(0, 1) \quad \text{and} \quad \theta \sim U[-\pi/2, \pi/2].$$

This means that

$$f_{X,\theta}(x,\theta) = 1/\pi, \quad 0 \leq x \leq 1, -\pi/2 \leq \theta \leq \pi/2$$

For the needle to intersect a line, we need $X < L \cos(\theta)$.

Expectations of Independent RVs (Theorem 4.2.10) Let X and Y be independent rvs.

- For any $A \subset \mathbb{R}$ and $B \subset \mathbb{R}$,

$$\Pr(X \in A, Y \in B) = \Pr(X \in A) \Pr(Y \in B)$$

i.e. the events $\{X \in A\}$ and $\{Y \in B\}$ are independent.

- Let $g(x)$ be a function only of x and $h(y)$ be a function only of y . Then

$$E[g(X)h(Y)] = [Eg(X)][Eh(Y)]$$

Proof:

$$\begin{aligned}
E[g(X)h(Y)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_{XY}(x,y)dx dy \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x)h(y)f_X(x)f_Y(y)dx dy \\
&= \left(\int_{-\infty}^{\infty} g(x)f_X(x)dx \right) \left(\int_{-\infty}^{\infty} h(y)f_Y(y)dy \right) \\
&= [Eg(X)][Eh(Y)]
\end{aligned}$$

Example X, Y are independent

$$\begin{aligned}
E(X^2Y^3) &= (EX^2)(EY^3) \\
E(Y^2Y^3) &\neq (EY^2)(EY^3)
\end{aligned}$$

Bivariate Transformation

Functions of random variables Let (X, Y) be a bivariate rv with known distributions. Define (U, V) by

$$U = g_1(X, Y), \quad V = g_2(X, Y)$$

Probability mapping For any Borel set $B \subset \mathbb{R}^2$,

$$\Pr[(U, V) \in B] = \Pr[(X, Y) \in A]$$

where A is the inverse mapping of B , i.e.

$$A = \{(x, y) \in \mathbb{R}^2 : (g_1(x, y), g_2(x, y)) \in B\}$$

The inverse is well defined even if the mapping is not bijective.

Example Let $g_1(x, y) = x, g_2(x, y) = x^2 + y^2$.

Discrete RVs Suppose that (X, Y) is a discrete rv, i.e. the pmf is positive on a countable set \mathcal{A} . Then (U, V) is also discrete and takes values on a countable set \mathcal{B} . Define

$$A_{u,v} = \{(x, y) \in \mathcal{A} : g_1(x, y) = u, g_2(x, y) = v\}$$

Then

$$f_{UV}(u, v) = \Pr(U = u, V = v) = \sum_{(x,y) \in A_{u,v}} f_{XY}(x, y)$$

Sum of two independent Poissons Let $X \sim \text{Poisson}(\lambda_1), Y \sim \text{Poisson}(\lambda_2)$, independent, and define

$$U = X + Y, \quad V = Y$$

- (X, Y) takes values in $\mathcal{A} = \{0, 1, 2, \dots\}^2$

- (U, V) takes values on $\mathcal{B} = \{(u, v) : v = 0, 1, 2, \dots, u = v, v + 1, v + 2, \dots\}$.
- For a particular (u, v) , $A_{uv} = \{(x, y) \in \mathcal{A} : x + y = u, y = v\} = (u - v, u)$.

The joint pmf of U and V is

$$f_{UV}(u, v) = f_{XY}(u - v, v) = \frac{e^{-\lambda_1} \lambda_1^{u-v}}{(u - v)!} \frac{e^{-\lambda_2} \lambda_2^v}{(v)!}$$

The distribution of $U = X + Y$ is the marginal

$$\begin{aligned} f_U(u) &= \sum_{v=0}^u \frac{e^{-\lambda_1} \lambda_1^{u-v}}{(u - v)!} \frac{e^{-\lambda_2} \lambda_2^v}{(v)!} \\ &= \frac{e^{-(\lambda_1 + \lambda_2)}}{u!} \sum_{v=0}^u \binom{u}{v} \lambda_1^{u-v} \lambda_2^v \\ &= \frac{e^{-(\lambda_1 + \lambda_2)}}{u!} (\lambda_1 + \lambda_2)^u \end{aligned}$$

We obtain that U is Poisson with parameter $\lambda = \lambda_1 + \lambda_2$.

Bivariate Transformations of Continuous RVs Suppose (X, Y) is continuous and the joint transformation

$$u = g_1(x, y), \quad v = g_2(x, y)$$

is one-to-one and differentiable. Define the inverse mapping

$$x = h_1(u, v), \quad y = h_2(u, v)$$

Then

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v)) |J(u, v)|$$

where $J(u, v)$ is the Jacobian of the transformation $(x, y) \rightarrow (u, v)$ given by

$$J(u, v) = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

Example: Rotation of a bivariate normal vector Let $X \sim N(0, 1)$, $Y \sim N(0, 1)$, independent. Define the rotation

$$\begin{aligned} U &= X \cos \theta - Y \sin \theta \\ V &= X \sin \theta + Y \cos \theta \end{aligned}$$

for fixed θ . Then $U \sim N(0, 1)$, $V \sim N(0, 1)$, independent.

Proof:

The range of (X, Y) is \mathbb{R}^2 . The range of (U, V) is \mathbb{R}^2 . Need the inverse transformation

$$\begin{aligned} X &= U \cos \theta + V \sin \theta \\ Y &= -U \sin \theta + V \cos \theta \end{aligned}$$

with Jacobian

$$J(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{vmatrix} = 1$$

The joint pdf of (X, Y) is

$$f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

The joint pdf of (U, V) is

$$\begin{aligned} f_{UV}(u, v) &= \frac{1}{2\pi} e^{-[(u \cos \theta + v \sin \theta)^2 + (-u \sin \theta + v \cos \theta)^2]/2} \cdot |1| \\ &= \frac{1}{2\pi} e^{-(u^2+v^2)/2} = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \end{aligned}$$

so $U \sim N(0, 1)$, $V \sim N(0, 1)$, and U and V are independent.

Functions of independent random variables (Theorem 4.3.5) Let X and Y be independent rvs. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ and $h : \mathbb{R} \rightarrow \mathbb{R}$ be functions. Then the random variables $U = g(X)$ and $V = h(Y)$ are independent.

Sum of two independent rvs Suppose X and Y are independent. What is the distribution of $Z = X + Y$? In general:

$$F_Z(z) = \Pr(X + Y \leq z) = \Pr(\{(x, y) \text{ such that } x + y \leq z\})$$

Various approaches:

- bivariate transformation method (continuous and discrete)
- Discrete convolution

$$f_Z(z) = \sum_{x+y=z} f_X(x) f_Y(y) = \sum_x f_X(x) f_Y(z-x)$$

- Continuous convolution (Section 5.2)
- MGF method (continuous and discrete)

Example (Sum of two independent Poissons) Define X, Y to be two independent random variables having Poisson distributions with parameters λ_i , $i = 1, 2$. Then:

$$f_{X,Y}(x, y) = \frac{e^{-\lambda_1} \lambda_1^x}{x!} \frac{e^{-\lambda_2} \lambda_2^y}{y!}, x, y = 0, 1, 2, \dots$$

The distribution of $S = X + Y$ is

$$\begin{aligned} f_S(s) &= \sum_{x=0}^s \frac{e^{-\lambda_1} \lambda_1^x}{x!} \frac{e^{-\lambda_2} \lambda_2^{s-x}}{(s-x)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{s!} \sum_{x=0}^s \binom{s}{x} \lambda_1^x \lambda_2^{s-x} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{s!} (\lambda_1 + \lambda_2)^s \end{aligned}$$

Again, S is Poisson with parameter $\lambda = \lambda_1 + \lambda_2$.

Moment generating function (Theorem 4.2.12) Let X and Y be independent rvs with mgfs $M_X(\cdot)$ and $M_Y(\cdot)$, respectively. Then the mgf of $Z = X + Y$ is

$$M_Z(t) = M_X(t)M_Y(t)$$

Proof:

$$\begin{aligned} M_Z(t) &= E \exp(Zt) &&= E\{\exp[(X + Y)t]\} \\ &= E[\exp(Xt) \exp(Yt)] &&= E[\exp(Xt)] \cdot E[\exp(Yt)] \\ &= M_X(t)M_Y(t) \end{aligned}$$

Corollary: If X and Y are independent and $Z = X - Y$,

$$M_Z(t) = M_X(t)M_Y(-t)$$

Example (sum of two independent Poissons) Suppose $X \sim \text{Poisson}(\lambda_X)$ and $Y \sim \text{Poisson}(\lambda_Y)$ and put $Z = X + Y$. Then, $Z \sim \text{Poisson}(\lambda_X + \lambda_Y)$. *Proof:*

$$\begin{aligned} M_Z(t) &= \exp[\lambda_X(e^t - 1)] \exp[\lambda_Y(e^t - 1)] \\ &= \exp[(\lambda_X + \lambda_Y)(e^t - 1)] \end{aligned}$$

Example (sum of two independent normals) Suppose $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ and X and Y are independent and $Z = X + Y$. Then

$$Z \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

Proof:

$$\begin{aligned} M_Z(t) &= \exp\left(\mu_x t + \frac{1}{2}\sigma_x^2 t^2\right) \exp\left(\mu_y t + \frac{1}{2}\sigma_y^2 t^2\right) \\ &= \exp\left[(\mu_x + \mu_y)t + \frac{1}{2}(\sigma_x^2 + \sigma_y^2)t^2\right] \end{aligned}$$

Example (sum of two independent gammas) Suppose $X \sim \Gamma(\alpha_x, \beta)$ and independently $Y \sim \Gamma(\alpha_y, \beta)$. Let $Z = X + Y$. Then $Z \sim \Gamma((\alpha_x + \alpha_y), \beta)$.

Proof:

$$\begin{aligned} M_Z(t) &= \left(\frac{1}{1 - \beta t} \right)^{\alpha_x} \left(\frac{1}{1 - \beta t} \right)^{\alpha_y} \\ &= \left(\frac{1}{1 - \beta t} \right)^{\alpha_x + \alpha_y} \end{aligned}$$

Remember that

- If $\alpha = 1$ we have an exponential with parameter β .
- If $\alpha = n/2$ and $\beta = 2$, we have a $\chi^2(n)$ (with n d.f.). The above result states that $\chi^2(n_1) + \chi^2(n_2) = \chi^2(n_1 + n_2)$.

Covariance and Correlation Let X and Y be two random variables with respective means μ_X, μ_Y and variances $\sigma_X^2 > 0$ and $\sigma_Y^2 > 0$, all assumed to exist.

- The *covariance* of X and Y is

$$\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}$$

- The *correlation* between X and Y is

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

also written as

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = E \left[\left(\frac{X - \mu_X}{\sigma_X} \right) \left(\frac{Y - \mu_Y}{\sigma_Y} \right) \right]$$

Properties Let c be a constant:

1. $\text{Cov}(X, X) = \text{Var}(X),$ $\text{Cor}(X, X) = 1$
2. $\text{Cov}(X, Y) = \text{Cov}(Y, X),$ $\text{Cor}(X, Y) = \text{Cor}(Y, X)$
3. $\text{Cov}(X, c) = 0,$ $\text{Cor}(X, c) = 0$
4. $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$

5. Let $X_c = X - \mu_X, Y_c = Y - \mu_Y$. Then

$$\begin{aligned} \text{Cov}(X, Y) &= \text{Cov}(X_c, Y_c) = E(X_c Y_c) \\ \text{Cor}(X, Y) &= \text{Cor}(X_c, Y_c) \end{aligned}$$

6. Let $\tilde{X} = (X - \mu_X)/\sigma_X, \tilde{Y} = (Y - \mu_Y)/\sigma_Y$. Then,

$$\text{Cor}(X, Y) = \text{Cor}(\tilde{X}, \tilde{Y}) = \text{Cov}(\tilde{X}, \tilde{Y}) = E(\tilde{X}\tilde{Y})$$

Independent vs. Uncorrelated

- X and Y are called *uncorrelated* iff

$$\text{Cov}(X, Y) = 0 \quad \text{or equivalently} \quad \rho_{XY} = 0$$

- If X and Y are independent and $\text{Cov}(X, Y)$ exists, then $\text{Cov}(X, Y) = 0$.
- If X and Y are uncorrelated, this does **not** imply that they are independent.

Example $X \sim U[-1, 1]$, $Y = X^2$. Then $\text{Cov}(X, Y) = 0$ but X, Y are not independent.

Correlation coefficient For any random variables X and Y ,

1. $-1 \leq \rho_{XY} \leq 1$
2. $|\rho_{XY}| = 1$ if and only if $\exists a \neq 0$ and b such that

$$\Pr(Y = aX + b) = 1.$$

if $\rho_{XY} = 1$ then $a > 0$, and if $\rho_{XY} = -1$, then $a < 0$.

proof:

Let $\tilde{X} = (X - \mu_X)/\sigma_X$, $\tilde{Y} = (Y - \mu_Y)/\sigma_Y$. Then $\text{Cor}(X, Y) = E(\tilde{X}\tilde{Y})$,

$$\begin{aligned} 1. \quad & 0 \leq E(\tilde{X} - \tilde{Y})^2 = 1 + 1 - 2E(\tilde{X}\tilde{Y}) \Rightarrow E(\tilde{X}\tilde{Y}) \leq 1 \\ & 0 \leq E(\tilde{X} + \tilde{Y})^2 = 1 + 1 + 2E(\tilde{X}\tilde{Y}) \Rightarrow -1 \leq E(\tilde{X}\tilde{Y}) \end{aligned}$$

$$\begin{aligned} 2. \quad & \rho_{XY} = 1 \iff \Pr(\tilde{Y} = \tilde{X}) = 1 \Rightarrow a > 0 \\ & \rho_{XY} = -1 \iff \Pr(\tilde{Y} = -\tilde{X}) = 1 \Rightarrow a < 0 \end{aligned}$$

Random Samples

Definition The random variables X_1, \dots, X_n are called a *random sample of size n from the population $f(x)$* if X_1, \dots, X_n are mutually independent and identically distributed (iid) random variables with the same pdf or pmf $f(x)$.

If X_1, \dots, X_n are iid, then their joint pdf or pmf is

$$f(x_1, \dots, x_n) = f(x_1)f(x_2) \dots f(x_n) = \prod_{j=1}^n f(x_j)$$

Statistics Let X_1, \dots, X_n be a random sample and let $T(x_1, \dots, x_n)$ be a function defined on \mathbb{R}^n . Then the random variable $Y = T(X_1, \dots, X_n)$ is called a *statistic*. The probability distribution of Y is called the *sampling distribution* of Y .

Note: T is only a function of (x_1, \dots, x_n) , no parameters.

Examples

$$\begin{aligned}\text{sample mean} \quad \bar{X} &= \frac{1}{n} \sum_{j=1}^n X_j \\ \text{sample variance} \quad S^2 &= \frac{1}{n-1} \sum_{j=1}^n (X_j - \bar{X})^2 \\ \text{sample standard deviation} \quad S &= \sqrt{S^2} \\ \text{minimum} \quad X_{(1)} &= \min_{1 \leq i \leq n} X_i\end{aligned}$$

Properties Let x_1, \dots, x_n be n numbers and define

$$\bar{x} = \frac{1}{n} \sum_{j=1}^n x_j, \quad s^2 = \frac{1}{n-1} \sum_{j=1}^n (x_j - \bar{x})^2$$

Then

$$\begin{aligned}\min_a \sum_{j=1}^n (x_j - a)^2 &= \sum_{j=1}^n (x_j - \bar{x})^2 \\ (n-1)s^2 &= \sum_{j=1}^n (x_j - \bar{x})^2 = \sum_{j=1}^n x_j^2 - n\bar{x}^2\end{aligned}$$

Residuals Lemma: Let X_1, \dots, X_n be a random sample from a population with mean μ and variance σ^2 . Define the residuals $R_i = X_i - \bar{X}$. Then

$$\begin{aligned}E(R_i) &= 0, \quad Var(R_i) = \frac{n-1}{n} \sigma^2 \\ Cov(R_i, \bar{X}) &= 0, \quad Cov(R_i, R_j) = -\sigma^2/n \text{ if } i \neq j\end{aligned}$$

Theorem Let X_1, \dots, X_n be a random sample from a population with mgf $M_X(t)$. Then the mgf of the sample mean is

$$M_{\bar{X}}(t) = [M_X(t/n)]^n$$

Convergence

Convergence in Probability A sequence of random variables X_1, \dots, X_n *converges in probability* to a random variable X , denoted

$$X_n \xrightarrow{p} X$$

if for every $\epsilon > 0$,

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| < \epsilon) = 1$$

or equivalently

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| > \epsilon) = 0$$

In other words, X_n is more and more likely to be close to X , or less and less likely to be far from X .

Example Let $X_n = X + \epsilon_n$, where $\epsilon_n \sim N(0, 1/n)$ and X is an arbitrary random variable. Then, as $n \rightarrow \infty$,

$$X_n \xrightarrow{p} X$$

Weak law of large numbers (WLLN) Let Y_1, \dots, Y_n be iid with common mean μ and variance σ^2 . Then, as $n \rightarrow \infty$,

$$\bar{Y}_n = \frac{1}{n} \sum_{j=1}^n Y_j \xrightarrow{p} \mu$$

Proof:

The proof is quite simple, being a straightforward application of Chebychev's Inequality. We have, for every $\epsilon > 0$,

$$\Pr(|\bar{Y}_n - \mu| \geq \epsilon) = \Pr(|\bar{Y}_n - \mu|^2 \geq \epsilon^2) \leq \frac{E(\bar{Y}_n - \mu)^2}{\epsilon^2} = \frac{Var(\bar{Y})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \text{ as } n \rightarrow \infty$$

Convergence in Distribution A sequence of random variables X_1, \dots, X_n *converges in distribution* to a random variable X , denoted

$$X_n \xrightarrow{d} X$$

if

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

This is also called *convergence in law* or *weak convergence*. In other words, the distribution of X_n is closer and closer to the distribution of X .

Relation between “in distribution” and “in probability” Theorem:

1. Convergence in probability implies convergence in distribution:

$$X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$$

2. Suppose $X_n \xrightarrow{d} X$ where X has a degenerate distribution, i.e. $\Pr\{X = a\} = 1$ for some $a \in \mathbb{R}$. Then,

$$X_n \xrightarrow{d} a \Rightarrow X_n \xrightarrow{p} a$$

Convergence in Distribution via Convergence of Mgf's Theorem: Suppose the mgf $M_n(t)$ of Y_n exists for $|t| < h$, and the mgf $M(t)$ of Y exists for $|t| < h_1 < h$. Then,

$$Y_n \xrightarrow{d} Y \iff \lim_{n \rightarrow \infty} M_n(t) = M(t), \quad |t| < h_1$$

Example Let $X_\lambda \sim \text{Poisson}(\lambda)$. Then, as $\lambda \rightarrow \infty$,

$$\begin{aligned} \frac{X_\lambda - \lambda}{\lambda} &\xrightarrow{p} 0 \\ \frac{X_\lambda - \lambda}{\sqrt{\lambda}} &\xrightarrow{d} N(0, 1) \end{aligned}$$

Central Limit Theorem Let X_1, X_2, \dots, X_n be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive $h > 0$). Let $EX_i = \mu$ and $Var(X_i) = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists) Define $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x , $-\infty < x < \infty$,

$$\lim_{n \rightarrow \infty} G_n(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution, in other words, $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0, 1)$

Proof:

Define $Y_i = (X_i - \mu)/\sigma$, and let $M_Y(t)$ denote the common mgf of Y_i s, which exists for $|t| < \sigma h$ and $M_Y(t) = M_{\frac{1}{\sigma}X_i - \mu/\sigma}(t) = e^{-\frac{\mu}{\sigma}t} M_X(\frac{t}{\sigma})$. Since

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i,$$

we have,

$$\begin{aligned} M_{\sqrt{n}(\bar{X}_n - \mu)/\sigma}(t) &= M_{\sum_{i=1}^n Y_i/\sqrt{n}}(t) \\ &= M_{\sum_{i=1}^n Y_i}(t/\sqrt{n}) \\ &= [M_Y(t/\sqrt{n})]^n. \end{aligned}$$

We now expand $M_Y(t/\sqrt{n})$ in a Taylor series (power series) around 0.

$$M_Y\left(\frac{t}{\sqrt{n}}\right) = \sum_{k=0}^{\infty} M_Y^{(k)}(0) \frac{(t/\sqrt{n})^k}{k!},$$

where $M_Y^{(k)}(0) = (d^k/dt^k)M_Y(t)|_{t=0}$. Since the mgfs exist for $|t| < h$, the power series expansion is valid if $t < \sqrt{n}\sigma h$.

Using the facts that $M_Y^{(0)} = 1$, $M_Y^{(1)} = 0$, and $M_Y^{(2)} = 1$ (by construction, the mean and variance of Y are 0 and 1), we have

$$M_Y\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{(t/\sqrt{n})^2}{2!} + R_Y\left(\frac{t}{\sqrt{n}}\right),$$

where R_Y is the remainder term in the Taylor expansion such that

$$\lim_{n \rightarrow \infty} \frac{R_Y(t/\sqrt{n})}{(t/\sqrt{n})^2} = 0.$$

Therefore, for any fixed t , we can write

$$\begin{aligned} \lim_{n \rightarrow \infty} \left[M_Y\left(\frac{t}{\sqrt{n}}\right) \right]^n &= \lim_{n \rightarrow \infty} \left[1 + \frac{(t/\sqrt{n})^2}{2!} + R_Y\left(\frac{t}{\sqrt{n}}\right) \right]^n \\ &= \lim_{n \rightarrow \infty} \left[1 + \frac{1}{n} \left(\frac{t^2}{2} + n R_Y\left(\frac{t}{\sqrt{n}}\right) \right) \right]^n \\ &= e^{t^2/2} \end{aligned}$$