Lecture 18

(16 October 2023)

Functions of Random Variables

$Y = g(X)$, i.e., $Y(\omega) = g(X(\omega))$.

For the discrete case

$$P_Y(y) = \sum_{x : g(x) = y} P_X(x).$$

For the continuous case we follow the two-step procedure outlined below.

1) Calculate the CDF F_Y of Y using

$$F_Y(y) = P(g(X) \leq y) = \int_{\{x : g(x) \leq y\}} f_X(x)\, dx.$$

2) Differentiate the CDF to obtain the PDF of Y:

$$f_Y(y) = \frac{d}{dy} F_Y(y).$$

Linear Function of a RV:

$$Y = aX + b, \quad a \neq 0.$$

Discrete case –

$$P_Y(y) = P_X\left(\frac{y-b}{a}\right), \quad \text{for all } y.$$

Continuous case –

$$Y = aX + b, \quad a \neq 0$$

$$\Rightarrow \quad f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Proof. $a > 0$ (proved earlier)

Let $a < 0$.

$$P(aX + b \leq y) = P\left(X \geq \frac{y-b}{a}\right)$$

$$= 1 - P\left(X < \frac{y-b}{a}\right)$$

$$= 1 - F_X\left(\frac{y-b}{a}\right)$$

$$\Rightarrow f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right).$$

Example. X is uniform on $[0, 1]$ and let $Y = \sqrt{X}$.

For $0 \leq y \leq 1$,

$$F_Y(y) = P(\sqrt{X} \le y)$$

$$= P(X \le y^2)$$

$$= y^2$$

$$\Rightarrow f_Y(y) = 2y, \quad 0 \le y \le 1.$$

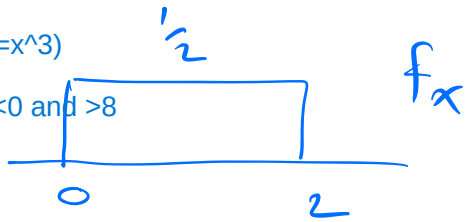
$F_Y(y) = 0$ $y \le 0$ and $F_Y(y) = 1$ $y \ge 1$.

$\therefore f_Y(y) = 0$ for y outside $[0,1]$.

Example. $Y = X^3$ $X \sim$ Uniform $[0\ 2]$.

if x can vary from 0 to 2, then realize that y can vary from 0 to 8 (because y=x^3)

Accordingly make the intervals. [0,8], <0 and >8

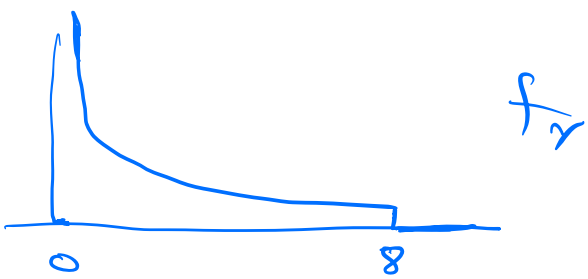


For $0 \le y \le 8$

$$F_Y(y) = P(X^3 \le y)$$

$$= P(X \le y^{1/3})$$

$$= \frac{y^{1/3}}{2}$$



$$\Rightarrow f_Y(y) = \frac{1}{6} y^{-2/3}, \quad 0 \le y \le 8.$$

Example. $Y = \frac{10}{X}$, $f_X(x) = \frac{1}{5}$, $5 \le x \le 10$.

For $1 \le y \le 2$

$$F_Y(y) = P\left(\frac{10}{X} \le y\right) = P\left(X \ge \frac{10}{y}\right)$$

$$= 1 - P\left(X \le \frac{10}{y}\right)$$

$$= 1 - \frac{1}{5}\left(-5 + \frac{10}{y}\right) = -\frac{2}{y} + 2$$

$$\Rightarrow f_Y(y) = \frac{+2}{y^2}, \quad 1 \le y \le 2.$$

Monotonic Function:

Let X be a continuous RV and suppose that its range is contained in a certain interval I, in the sense that $f_X(x) = 0$ for $x \notin I$. We consider $Y = g(X)$, and assume that g is strictly monotonic over the interval I:

X is just a random variable.

$g(x) < g(x')$ for all x, x' s.t. $x < x'$

(monotonically increasing)

or

$g(x) > g(x')$ for all x, x' s.t. $x < x'$.

(monotonically decreasing)

For $y \in g^{-1}(I)$,

$$P(g(X) \le y)$$

$$\overset{\text{incr. } g}{=} P(X \le g^{-1}(y)) = F_X(g^{-1}(y))$$

$$\Rightarrow f_Y(y) = f_X(g^{-1}(y))\,(g^{-1})'(y).$$

For decreasing g

$$F_Y(y) = P(g(x) \le y) = P(x \ge g^{-1}(y))$$

$$= 1 - P(x \le g^{-1}(y)) = 1 - F_x(g^{-1}(y))$$

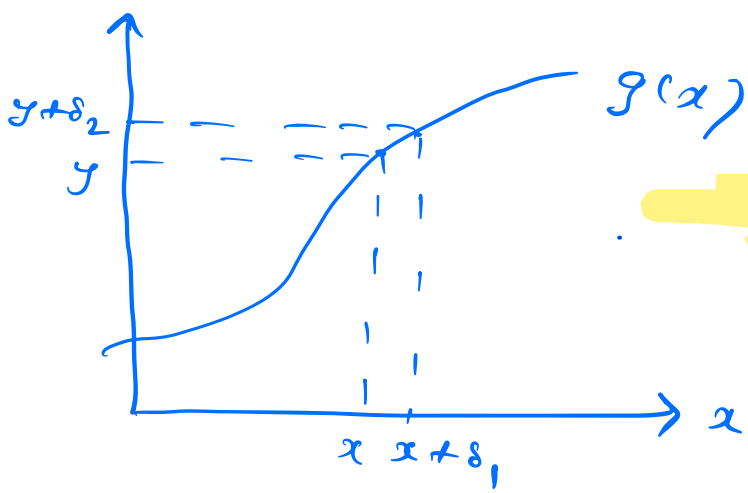
$$\Rightarrow f_Y(y) = -f_x(g^{-1}(y))\,(g^{-1})'(y).$$

$$\therefore f_Y(y) = f_x(g^{-1}(y))\,|(g^{-1})'(y)|.$$

This can also be written as

$$f_Y(y) = \begin{cases} f_x(g^{-1}(y)) / |g'(g^{-1}(y))| & \text{if } g(x) = y \text{ for some } x \text{ s.t. } f_x(x) > 0 \\ 0 & \text{if } g(x) \neq y \text{ for all } x \text{ with } f_x(x) > 0 \end{cases}$$

Illustration.



$\delta_1, \delta_2 \approx 0$

$$\frac{\delta_2}{\delta_1} = g'(x)$$

$$P(y \le Y \le y+\delta_2) \approx f_Y(y)\,\delta_2$$

$$= P(x \le X \le x+\delta_1) = f_x(x)\,\delta_1$$

why are both of them equal?

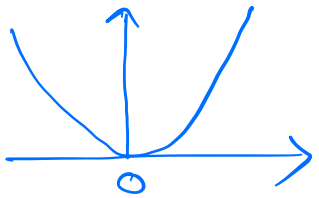
because P(X=x) = P(Y=y) because there is only one point since one to one function. and small delta doesn't make a difference.

treat fY(y) as a constant in this interval

$$f_Y(y) = f_X(x)\frac{\delta_1}{\delta_2} = \frac{f_X(x)}{g'(x)} = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}.$$

Non-monotonic function

$Y = g(X)$, $g(x) = x^2$



For $y \geq 0$

$$P(g(x) \leq y) = P(x^2 \leq y)$$

$$= P(-\sqrt{y} \leq x \leq \sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

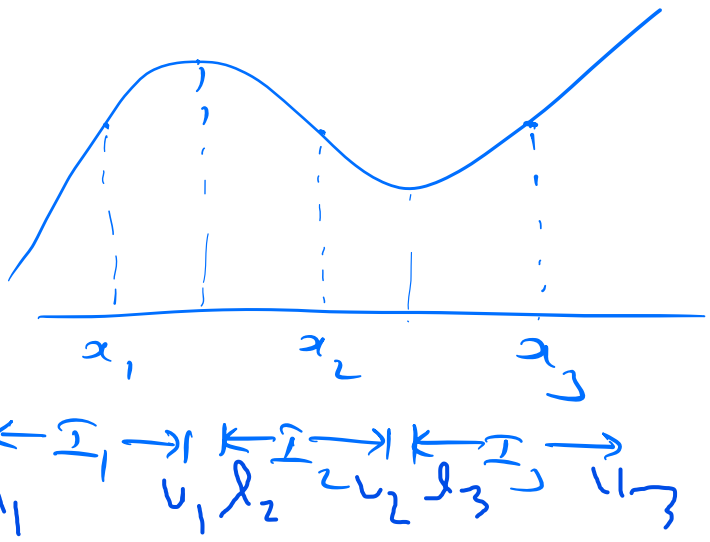
$$\Rightarrow f_Y(y) = \frac{f_X(\sqrt{y})}{2\sqrt{y}} + \frac{f_X(-\sqrt{y})}{2\sqrt{y}}.$$

Theorem. Consider a continuous RV X with PDF f_X and let $Y = g(X)$. Suppose we partition the domain of f_X into a finite number of intervals such that $g(x)$ is strictly monotone and differentiable on each interval. Then the PDF of Y is given by

$$f_y(y) = \sum_{i=1}^{n} \frac{f_x(x_i)}{|g'(x_i)|} \quad \text{where}$$

$x_1, x_2, \ldots, x_n$ are real solutions to $g(x) = y$ in the domain of f_x.

Proof.



$I_1 = [l_1, u_1]$

$I_2 = [l_2, u_2]$

$I_3 = [l_3, u_3]$.

For a given y let $g(x_i) = y$ $i \in \{1, 2, 3\}$.

$g(x)$ is increasing at x_1

decreasing at x_2 and

increasing at x_3.

$$P(g(x) \le y) = P(\{l_1 \le x \le x_1\} \cup \{x_2 \le x \le u_2\} \cup \{l_3 \le x \le x_3\})$$

$$= P(l_1 \le x \le x_1) + P(x_2 \le x \le u_2) + P(l_3 \le x \le x_3)$$

$$\Rightarrow f_y(y) = f_x(x_1)\frac{d}{dy}x_1 - f_x(x_2)\frac{dx_2}{dy} + f_x(x_3)\frac{d}{dy}x_3$$

$$= \frac{f_x(x_1)}{|g'(x_1)|} + \frac{f_x(x_2)}{|g'(x_2)|} + \frac{f_x(x_3)}{|g'(x_3)|}.$$

→ For $y = x^2$, $x_1 = \sqrt{y}$, $x_2 = -\sqrt{y}$.

So $f_y(y) = \frac{f_x(\sqrt{y})}{2\sqrt{y}} + \frac{f_x(-\sqrt{y})}{2\sqrt{y}}$.

Lecture 19

(26 October 2023)

Functions of Two Random Variables

$$Z = g(X, Y),$$

i.e., $Z(\omega) = g(X(\omega), Y(\omega))$.

Example. $Z = \max(X, Y)$, X & Y are independent.

$$P(Z \le z) = P(\max(X, Y) \le z)$$

$$= P(X \le z, Y \le z)$$

$$= P(X \le z, Y \le z)$$

$$= F_X(z) F_Y(z)$$

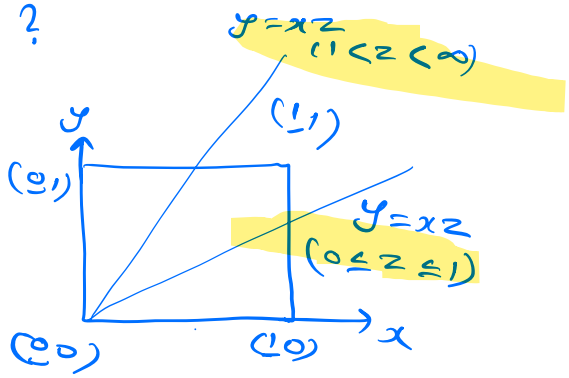
$$\Rightarrow f_Z(z) = F_X(z) f_Y(z) + f_X(z) F_Y(z).$$

As a special case $X, Y \sim$ Uniform $[0, 1]$.

$$f_Z(z) = \begin{cases} 2z, & 0 \le z \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Exercise: $Z = \min\{X, Y\}$, X and Y are indep. Find f_Z.

Example. X and Y are independent RVs that are uniformly distributed on $[0,1]$. What is the PDF of $Z = Y/X$?



$$P(Z \le z) = P(Y/X \le z)$$

z can be from 0 to infinity

For $0 \le z \le 1$

$$P(Y \le Xz) = \int_{x=0}^{1} \int_{y=0}^{zx} 1 \, dx \, dy = z/2.$$

For $1 < z < \infty$

$$P(Y \le Xz) = 1 - \frac{1}{2z}.$$

$$\Rightarrow f_Z(z) = \begin{cases} z/2 & \text{if } 0 \le z \le 1 \\ 1 - 1/2z & \text{if } z > 1 \\ 0 & \text{o.w.} \end{cases}$$

$$f_Z(z) = \begin{cases} 1/2 & \text{if } 0 \le z \le 1 \\ 1/2z^2 & \text{if } z > 1 \\ 0 & \text{o.w.} \end{cases}$$

Sum of Independent Random Variables

Let $Z = X + Y$ where X and Y are independent random variables.

Discrete case:

$$P_Z(z) = P(X+Y=z)$$

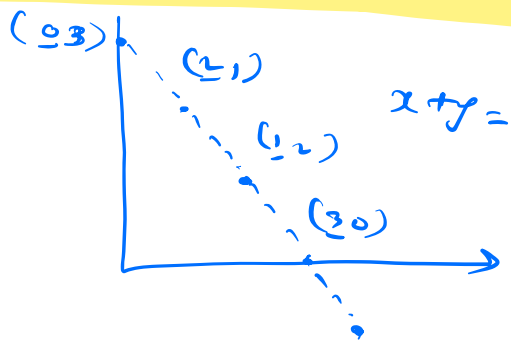
$$= \sum_{(x,y):\, x+y=z} P_{XY}(x,y)$$

$$= \sum_x P_{XY}(x, z-x) = \sum_x P_X(x)\, P_Y(z-x).$$

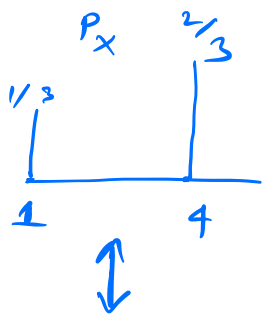
because X and Y are independent RVs

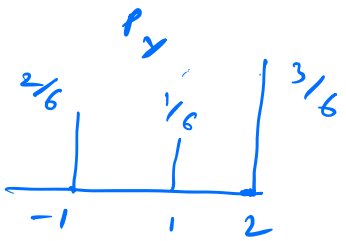
The resulting PMF P_Z is called the convolution of the PMFs of X and Y.

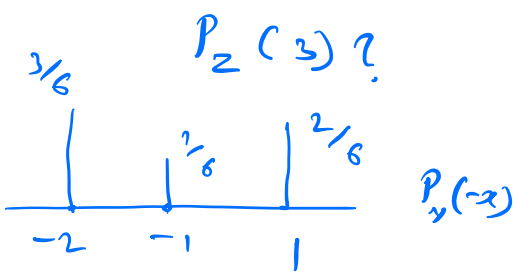
$z = 3 \Rightarrow$

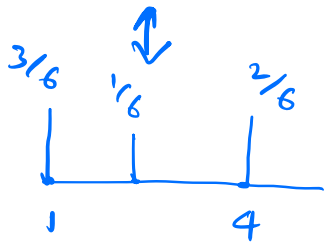


Sum of joint PMF at all points on line $x+y=3$.









$P_Y(3-x)$

$$\Rightarrow \sum_x P_X(x)\, P_Y(3-x) = P_Z(3).$$

Continuous case:

$$P(X+Y \le z) = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{z-x} f_{XY}(x,y)\, dy\, dx$$

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

$$= \int_{x=-\infty}^{\infty} f_X(x)\, F_Y(z-x)\, dx$$

$$\Rightarrow f_Z(z) = \frac{d}{dz} F_Z(z)$$

$$= \int_{-\infty}^{\infty} f_X(x)\, \frac{d}{dz} F_Y(z-x)\, dx$$

$$\frac{d}{dz}\left(F_Y(z-x)\right) = \frac{d}{dy}\left(F_Y(z-x)\right) \cdot \frac{dy}{dz}$$

$$= \int_{x=-\infty}^{\infty} f_X(x)\, f_Y(z-x)\, dx$$

$y + x = z$

$f_Y(z-x)$ chn. $\rightarrow \frac{dy}{dz} = 1$

Theorem. If $X \sim N(0, 1)$, $Y \sim N(0, 1)$ are independent then

$$X + Y \sim N(0, 2).$$

Proof. $Z = X + Y$.

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2 - z^2 - x^2 + 2zx}{2}} \cdot dx$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2 - zx + z^2}{2}} dx$$

$$= \frac{1}{\sqrt{4\pi}} \cdot e^{-z^2/4} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi}} e^{-(x - z/2)^2} dx$$

$$= \frac{1}{\sqrt{4\pi}} \cdot e^{-z^2/4} \cdot \underbrace{\qquad\qquad}_{= 1}$$

Exercise. $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$ indep $\Rightarrow X + Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$.

Sum of a discrete RV and a continuous RV:

$$Z = X + Y$$

X - discrete Y - continuous X & Y are independent

$$P(Z \le z) = P(X+Y \le z)$$

$$= \sum_x P(X+Y \le z \mid X=x)\, P_X(x)$$

$$= \sum_x P(Y \le z-x \mid X=x)\, P_X(x)$$

$$= \sum_x P(Y \le z-x)\, P_X(x)$$

$$= \sum_x F_Y(z-x)\, P_X(x)$$

$$= \sum_x \int_{-\infty}^{z-x} f_Y(y)\, dy\, P_X(x) \qquad y = t-x$$

$$= \sum_x \int_{-\infty}^{z} f_Y(t-x)\, dt\, P_X(x)$$

$$= \int_{-\infty}^{z} \left(\sum_x f_Y(t-x)\, P_X(x) \right) dt$$

$$\therefore f_Z(z) = \sum_x f_Y(z-x)\, P_X(x).$$

Functions of Two Random Variables

$(X, Y) \sim f_{XY}$.

$Z = g_1(X, Y)$, $W = g_2(X, Y)$.

$$P(Z \le z, W \le w) = P(g_1(X,Y) \le z, g_2(X,Y) \le w)$$

$$= \iint_{\substack{x,y: \\ g_1(x,y) \le z \\ g_2(x,y) \le w}} f_{XY}(x, y)\, dx\, dy$$

Sometimes it may be easier to find out the region under the above integral and compute the integral. Then we can find f_{ZW} by taking double derivative:

$$f_{ZW}(z, w) = \frac{\partial^2 F_{ZW}(z, w)}{\partial z\, \partial w}.$$

Exercise. Suppose x and y are independent uniformly distributed RVs in $[0,1)$.

$Z = \min\{x,y\}$ $W = \max\{x,y\}$. Find f_{ZW}.

- In some cases we may be able to obtain a closed-form expression for f_{ZW} in terms of f_{xy}.

Assume that g_1 & g_2 satisfy:

(1) $z = g_1(x,y)$, $w = g_2(x,y)$

$\Leftrightarrow$ $x = h_1(z,w)$ $y = h_2(z,w)$.

Example: $z = x+y$ $w = x-y$.

(2) The functions g_1 & g_2 have continuous partial derivatives at all points (x,y) and are such that

$$J(x,y) = \begin{vmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{vmatrix} = \frac{\partial g_1}{\partial x}\frac{\partial g_2}{\partial y} - \frac{\partial g_1}{\partial y}\frac{\partial g_2}{\partial x} \neq 0$$

at all points (x,y).

Theorem. In the setting above,

$$f_{ZW}(z,w) = f_{XY}(x,y)\,|J(x,y)|^{-1}.$$

This is analogous to

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|},$$

when $Y = g(X)$ for a monotonic g.

Lecture 20

(30 October 2023)

Two Functions of Two Random Variables

$(X, Y) \sim f_{XY}$.

Let $Z = g_1(X, Y)$, $W = g_2(X, Y)$.

Suppose g_1 & g_2 are such that

(i) $\exists$ functions h_1, h_2 satisfying

$$x = h_1(Z, W), \quad y = h_2(Z, W).$$

(ii)
$$J(x, y) := \begin{vmatrix} \frac{\partial g_1(x,y)}{\partial x} & \frac{\partial g_1(x,y)}{\partial y} \\ \frac{\partial g_2(x,y)}{\partial x} & \frac{\partial g_2(x,y)}{\partial g_2} \end{vmatrix}$$

$$= \frac{\partial g_1}{\partial x}\frac{\partial g_2}{\partial y} - \frac{\partial g_2}{\partial x}\frac{\partial g_1}{\partial y} \neq 0$$

at all x, y.

Theorem. In the setting above,

$$f_{ZW}(z,w) = f_{XY}(x,y)\,|J(x,y)|^{-1},$$

where $x = h_1(z,w)$ $y = h_2(z,w)$.

This is analogous to

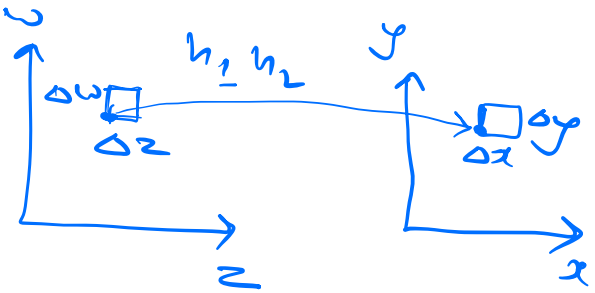
$$f_Y(y) = \frac{f_X(g^{-1}(y))}{|g'(g^{-1}(y))|},$$

when $Y = g(X)$ for a monotonic g.

Proof. Consider

$$P(z \le Z \le z+\delta z,\ w \le W \le w+\delta w)$$

$$= f_{ZW}(z,w)\,\delta z\,\delta w.$$



Alternately,

$$P(z \le Z \le z+\delta z,\ w \le W \le w+\delta w)$$

$$= P(x \le X \le x+\delta x,\ y \le Y \le y+\delta y) = f_{XY}(x,y)\,\delta x\,\delta y$$

where $x = h_1(z,w)$ $y = h_2(z,w)$.

$$\Rightarrow f_{ZW}(z,w) = f_{XY}(x,y)\,\frac{\Delta x\,\Delta y}{\Delta z\,\Delta w}$$

$$= \frac{f_{XY}(x,y)}{\left(\frac{\Delta z\,\Delta w}{\Delta x\,\Delta y}\right)} = \frac{f_{XY}(x,y)}{|J(x,y)|},$$

where

$$J(x,y) = \frac{\partial g_1}{\partial x}\frac{\partial g_2}{\partial y} - \frac{\partial g_1}{\partial y}\frac{\partial g_2}{\partial x}.$$

[See chapter 6 from Papoulis & Pillai's Book for more details and a formal justification]

Analogy: $y = g(x) \Rightarrow \frac{\Delta y}{\Delta x} = g'(x)$,

Similarly $z = g_1(x,y)$ $w = g_2(x,y)$

$$\Rightarrow \frac{\Delta z\,\Delta w}{\Delta x\,\Delta y} = \left|\det\begin{bmatrix} \partial g_1/\partial x & \partial g_1/\partial y \\ \partial g_2/\partial x & \partial g_2/\partial y \end{bmatrix}\right|.$$

- Note that Jacobian is essentially doing the following change of variables:

$$P((Z,W)\in B) = P((X,Y)\in A) = \iint_A f_{XY}(x,y)\,dx\,dy$$

$$= \iint_B f_{XY}(h_1(z,w), h_2(z,w))\cdot \frac{1}{|J(h_1(z,w), h_2(z,w)|}\,dz\,dw,$$

→ more generally for a given point (z, w) $g_1(x,y) = z$ $g_2(x,y) = w$ can have many solutions.

Let $(x_1,y_1)\ (x_2,y_2)\ \cdots\ (x_n,y_n)$ represent these multiple solutions such that

$$g_1(x_i,y_i) = z \quad g_2(x_i,y_i) = w \quad i \in [1:n].$$

Then

$$f_{zw}(z,w) = \sum_{i=1}^{n} f_{xy}(x_i,y_i)\,|J(x_i,y_i)|^{-1}.$$

Example. $\begin{bmatrix} z \\ w \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$.

$$J(x,y) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \neq 0.$$

$$f_{zw}(z,w) = \frac{f_{xy}(x,y)}{|a_{11}a_{22} - a_{12}a_{21}|}, \quad \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} \begin{bmatrix} z \\ w \end{bmatrix}.$$

Moment Generating Functions

— A transform associated with a random variable and provides an alternative representation of a probability law.

— Particularly useful for certain types of mathematical manipulations

— Moment generating function (MGF) of a RV X is a function $M_X(s)$ defined as

$$M_X(s) = E[e^{sX}], \text{ for a scalar } s.$$

Discrete case:

$$M_X(s) = \sum_x e^{sx} P_X(x).$$

Continuous case:

$$M_X(s) = \int_{-\infty}^{\infty} e^{sx} f_X(x)\, dx.$$

$M_X(0) = 1.$

Example. Let $P_X(1) = \frac{1}{2}$, $P_X(2) = \frac{1}{4}$, $P_X(3) = \frac{1}{4}$.

$$M_X(s) = \frac{1}{2}e^{s} + \frac{1}{4}e^{2s} + \frac{1}{4}e^{3s}.$$

Example. Exponential RV with parameter λ.

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0.$$

$$M_X(s) = \int_0^\infty e^{sx}\,\lambda e^{-\lambda x}\,dx$$

$$= \lambda \int_0^\infty e^{(s-\lambda)x}\,dx$$

$$= \lambda \left. \frac{e^{(s-\lambda)x}}{(s-\lambda)} \right|_0^\infty$$

$$= \frac{\lambda}{\lambda - s} \quad \text{if } s < \lambda,$$

otherwise the integral is infinite.

— $M_X(s)$ is only defined for those values of s for which $E[e^{sx}]$ is finite.

Exercise

1) If X takes only non-negative integer values then show that

$$\lim_{s \to -\infty} m_X(s) = P(X=0).$$

2) Let X be a Poisson RV with parameter λ. Find $m_X(s)$.

3) Prove that $m_{ax+b}(s) = e^{sb} m_X(as)$.

Example

$X \sim N(0,1)$.

$$m_X(s) = \int_{-\infty}^{\infty} e^{sx} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2} + sx} dx$$

$$= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-s)^2}{2}} \cdot e^{s^2/2} dx$$

$$= e^{s^2/2} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-(x-s)^2/2} dx$$

$$= e^{s^2/2}.$$

$Y \sim N(\mu, \sigma^2) \Rightarrow M_Y(s) = e^{sb} m_X(as)$

$[Y = \sigma X + \mu]$

$= e^{sb} \cdot e^{\sigma^2 s^2/2}$

$= e^{\mu s + \sigma^2 s^2/2}$.

- MGFs are useful mainly for two reasons

(1) MGFs of x gives us all moments of x.

$$M_x(s) = E[e^{sx}]$$

$$= \int_{-\infty}^{\infty} e^{sx} f_x(x)\, dx$$

$$\frac{d}{ds} m_x(s) = \int_{-\infty}^{\infty} x\, e^{sx} f_x(x)\, dx$$

$$\frac{d}{ds} m_x(s)\Big|_{s=0} = \int_{-\infty}^{\infty} x f_x(x)\, dx$$

Similarly $\frac{d^n}{ds^n} m_x(s)\Big|_{s=0} = \int_{-\infty}^{\infty} x^n f_x(x)\, dx.$

(2) MGF (if it exists) uniquely determines

the CDF of random variable.

Consider two random variables X and Y. Suppose that there exists a positive constant c such that MGFs of X and Y are finite and identical for all values of s in $[-c, c]$. Then, ⟸ Uniqueness property

$$F_X(t) = F_Y(t) \text{ for all } t \in R.$$

proof follows from connection with Laplace transform.

This can also be seen through characteristic functions.

$$\Phi_X(t) = E[e^{itx}] \quad i = \sqrt{-1}.$$

- characteristic functions always exist.
- Related to fourier transform.

Fact:

$\int_{-\infty}^{\infty} g(x)\,dx$ exists if $\int_{-\infty}^{\infty} |g(x)|\,dx$ exists.

since $|e^{itx}| = 1$ Φ_X always exists.

— If X and Y are independent RVs

$$M_{X+Y}(s) = M_X(s)\, M_Y(s).$$

Exercise. (i) $X \sim N(\mu_x, \sigma_x^2)$, $Y \sim N(\mu_y, \sigma_y^2)$

$\Rightarrow X+Y \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$.

(ii) Let X and Y be independent Poisson RVs with parameters λ_1 & λ_2, respectively. Then $X+Y$ is Poisson with parameter $\lambda_1 + \lambda_2$

Lecture 21

(2 November 2023)

Module 5 : Probability Bounds & Limit Theorems

- Suppose we want to compute $P(X \geq a)$. In some scenarios it may be sufficient to have bounds on this probability instead of its exact value, e.g., when the distribution of X is unavailable or hard to compute. In such scenarios if we have exact values or bounds for the mean and variance of X, we can obtain meaningful bounds on the quantity of interest.

Markov's Inequality

If X is a non-negative random variable then $P(X \geq a) \leq \frac{E[X]}{a}$, for all $a > 0$.

Interpretation:

"If $x \geq 0$ and $E[X]$ is small, then the probability that x takes a large value must be small".

$E[X] = \sum_x x P_X(x)$, $P_X(100) \to$ small

proof. we first prove that if $x \geq y$, then $E[X] \geq E[Y]$.

$z = x - y \geq 0$, i.e., $x(\omega) - y(\omega) \geq 0$, $\forall \omega \in \Omega$.

$E[z] \geq 0 \Rightarrow E[X] \geq E[Y]$.

Let $Y = a\,1\{X \geq a\}$. we have

$X \geq Y$.

$\Rightarrow E[X] \geq E[Y]$

$\Rightarrow E[X] \geq E[a\,1\{X \geq a\}]$

$= a\,P(X \geq a)$

$\therefore P(X \geq a) \leq \frac{E[X]}{a}$.

Example. Let $X \sim \text{Binomial}(n, p)$. Using Markov's inequality find an upper bound on $P(X \geq \alpha n)$ where $p < \alpha < 1$. Evaluate the bound for $p = 1/2$ and $\alpha = 3/4$.

$$P(X \geq \alpha n) \leq \frac{E[X]}{\alpha n} = \frac{np}{n\alpha} = p/\alpha$$

$$= 1/2 \cdot \frac{4}{3} = 2/3.$$

$$\therefore P(X \geq \alpha n) \leq \frac{2}{3}.$$

Example. $X \sim \text{Uniform}[-4, 4]$.

$$P(X \geq 3) \leq P(|X| \geq 3)$$

$$\leq \frac{E[|X|]}{3}$$

$$= \frac{2}{3}.$$

Chebyshev's Inequality

If X is a random variable with mean μ and variance σ^2 then

$$P(|X-\mu| \geq c) \leq \frac{\sigma^2}{c^2} \quad \text{for all } c>0.$$

Proof. Let $Y = |X-\mu|$.

$$P(Y \geq c) = P(Y^2 \geq c^2)$$

$$\leq \frac{E[Y^2]}{c^2} = \frac{E[|X-\mu|^2]}{c^2}$$

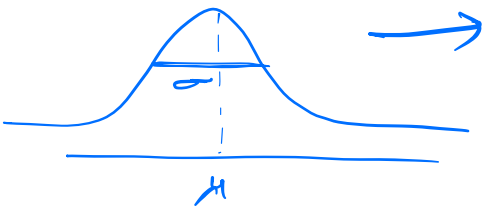
$$= \frac{\sigma_X^2}{c^2}.$$

Interpretation:

"If a random variable has small variance then the probability that it takes a value far from its mean is also small."

- Recall that variance measures the spread

of RV x around its mean.



→ small σ
⇒ dispersion around mean is small

- An alternative form of chebyshev's inequality:

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

Example. $X \sim$ Binomial (n, p). Using chebyshev's inequality find an upper bound on $P(X \geq \alpha n)$ where $p < \alpha < 1$. Evaluate for $p = \frac{1}{2}$, $\alpha = \frac{3}{4}$.

$$P(X \geq \alpha n) = P(X - np \geq n(\alpha - p))$$

$$\leq P(|X - np| \geq n(\alpha - p))$$

$$\leq \frac{\text{Var}(X)}{n^2(\alpha-p)^2} = \frac{np(1-p)}{n^2(\alpha-p)^2}$$

$$= \frac{4}{n}.$$

Chernoff Bounds

If X is a random variable then for any $a \in \mathbb{R}$ we can write

$$P(X \geq a) = P(e^{sX} \geq e^{sa}) \quad \text{for } s > 0$$

$$P(X \leq a) = P(e^{sX} \geq e^{sa}) \quad \text{for } s < 0.$$

now apply Markov Inequality here

Note that e^{sX} is always a positive random variable for all $s \in \mathbb{R}$. Thus we can apply Markov's inequality.

For $s > 0$ $\quad P(X \geq a) = P(e^{sX} \geq e^{sa})$

$$\leq \frac{E[e^{sX}]}{e^{sa}}$$

$$\leq \min_{s>0} \; E[e^{sX}] \cdot e^{-as}.$$

Similarly for $s < 0$

$$P(X \leq a) \leq \min_{s<0} \; E[e^{sX}] \, e^{-as}.$$

Example. $X \sim \text{Binomial}(n,p)$. Using chernoff bounds bound $P(X \geq \alpha n)$ where $p < \alpha < 1$. Evaluate the bound for $p = \frac{1}{2}$ and $\alpha = \frac{3}{4}$.

$$X = \sum_{i=1}^{n} X_i \qquad P_{X_i}(1) = p = 1 - P_{X_i}(0)$$

$X_i \sim$ i.i.d.

$$E[e^{sX}] = M_X(s)$$

$$= \prod_{i=1}^{n} M_{X_i}(s)$$

$$= (pe^s + 1-p)^n.$$

$$P(X \geq \alpha n) \leq \min_{s>0} e^{-\alpha n s} (pe^s + 1-p)^n$$

$$\frac{d}{ds}\left(e^{-\alpha n s}(pe^s + 1-p)^n\right) = 0$$

find the s that minimizes the function

$$\Rightarrow e^s = \frac{\alpha(1-p)}{p(1-\alpha)}$$

$$\Rightarrow s = \log \frac{\alpha(1-p)}{p(1-\alpha)} > 0$$

Since $\frac{1-p}{p} \cdot \frac{\alpha}{1-\alpha} > 1$,

Also check double derivative ≥ 0.

we get

$$P(X \geq \alpha n) \leq \left(\frac{\alpha(1-p)}{p(1-\alpha)}\right)^{-\alpha n} \left(\frac{p\alpha(1-p)}{p(1-\alpha)} + 1-p\right)^{n}$$

$$= \left(\frac{\alpha(1-p)}{p(1-\alpha)}\right)^{-\alpha n} \left(\frac{1-p}{1-\alpha}\right)^{n}$$

$$= \left(\frac{1-p}{1-\alpha}\right)^{(1-\alpha)n} \left(\frac{p}{\alpha}\right)^{\alpha n}$$

$$= \left(\frac{1-1/2}{1-3/4}\right)^{\frac{n}{4}} \cdot \left(\frac{2}{3}\right)^{3n/4}$$

$$= \frac{2^{n/4} \cdot 2^{3n/4}}{27^{n/4}} = \left(\frac{16}{27}\right)^{n/4}.$$

comparison between markov, chebyshev and chemoff bounds:

$P(X \geq \alpha n) \leq \frac{2}{3}$ [markov]

$P(X \geq \alpha n) \leq 4/n$ [chebyshev]

$P(X \geq \alpha n) \leq \left(\frac{16}{27}\right)^{n/4}$ [chemoff]

- The bound given by Markov is the weakest bound. It is a constant (does not depend on n).
- Chebyshev's bound is stronger than Markov's. In particular note that $4/n \to 0$ as $n \to \infty$.
- Chernoff bound is the strongest bound. It goes to zero exponentially fast.

Example. Suppose X is a RV taking values in $[a,b]$. Obtain a bound on $P(|X-\mu| \geq c)$ using Chebyshev's inequality.

claim: $X \in [a,b] \Rightarrow \sigma_X^2 \leq \frac{(b-a)^2}{4}$.

$E[(X-v)^2] = E[X^2] - 2E[X]v + v^2$

$\sigma_X^2 \leq E[(X - \frac{a+b}{2})^2] = E[((X-a)+(X-b))^2]/4$

$= \frac{1}{4} E[((X-a)-(X-b))^2 + 4(X-a)(X-b)]$

$\leq (b-a)^2/4$.

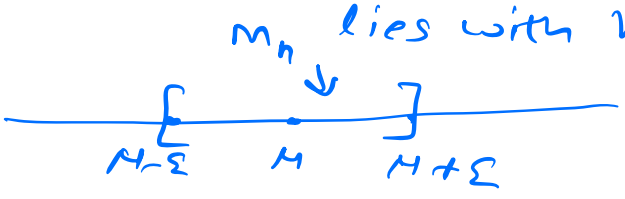
$P(|X-\mu| \geq c) \leq \frac{\sigma_X^2}{c^2} \leq \frac{(b-a)^2}{4c^2}$.

Weak Law of Large Numbers (WLLN)

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed (i.i.d.) random variables with mean μ. For every $\varepsilon > 0$, we have

$$P\left(\left| \frac{\sum_{i=1}^{n} X_i}{n} - \mu \right| \geq \varepsilon \right) \to 0 \quad \text{as} \quad n \to \infty.$$

Interpretation: Let $M_n = \sum_{i=1}^{n} X_i / n$. WLLN asserts that the sample mean of a large number of i.i.d. RVs is very close to the true mean.



Proof. $E[M_n] = \mu$, $\text{Var}(M_n) = \frac{\sigma^2}{n}$.

$$P(|M_n - E[M_n]| \geq \epsilon)$$

$$= P(|M_n - \mu| \geq \epsilon) \leq \frac{\mathrm{Var}(M_n)}{\epsilon^2}$$

$$= \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0 \quad \text{as} \quad n \to \infty,$$

Lecture 22

(6 November 2023)

WLLN. Let $x_1, x_2, \ldots$ be a sequence of independent and identically distributed RVs with mean μ. For every $\varepsilon > 0$ we have

$$P\left(\left| \frac{\sum_{i=1}^{n} x_i}{n} - \mu \right| \geq \varepsilon \right) \to 0 \quad \text{as} \quad n \to \infty.$$

Remark. Note that the proof we have seen assumes finite variance. It turns out that this law remains true even if the x_i have infinite variance, but a much more elaborate argument is needed.

Another Interpretation:

Let $a_1, a_2, \ldots, a_n$ be the realizations of a sequence of i.i.d. random variables $\sim p_X$.

$$a_i \in \mathcal{X} = \{x_1, x_2, \ldots, x_k\}.$$

$$\sum_{i=1}^{n} a_i / n = \sum_{i=1}^{k} x_i k_i / n \quad , \quad \sum_{i=1}^{k} k_i = n$$

k_i = no. of times x_i occurs

If we interpret probabilities as relative frequency

$$\sum_{i=1}^{k} \left(\frac{k_i}{n}\right) x_i \longrightarrow \sum_{i=1}^{k} p_X(x_i)\, x_i = E[X].$$

Example (Polling). Each voter in a population selects a particular candidate A with probability p and chooses another with probability $1-p$.

$$X_i = \begin{cases} 1 & i^{th} \text{ voter chooses A} \\ 0 & \text{o.w.} \end{cases}$$

$$p_{X_i}(1) = p = 1 - p_{X_i}(0).$$

X_i are independent and identically distributed.

We are interested in knowing the value of p.

$$M_n = \sum_{i=1}^{n} X_i / n.$$

the mean here is p.

$$P(|M_n - p| \geq 0.1) \leq \frac{p(1-p)}{n(0.1)^2} \leq \frac{1}{400(0.1)^2} \quad \text{for } n = 100$$

$$= 0.25$$

With a sample size of 100 the probability that our estimate is incorrect by more than 0.1 is smaller than 0.25.

$$P(|M_n - p| \geq 0.01) \leq \frac{1}{4n(0.01)^2}$$

$$n \geq 50000 \Rightarrow P(|M_n - p| \geq 0.01) \leq 0.05.$$

with a sample size of 100, the probability that our estimate is incorrect by more than 0.01 is smaller than 0.05

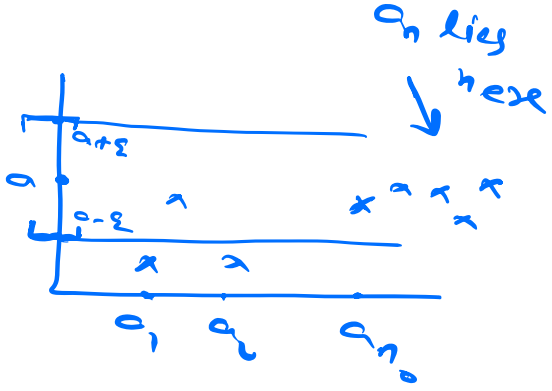
Convergence in Probability

We can interpret WLLN as stating that "$M_n = \sum_{i=1}^{n} X_i/n$ converges to mean M" in the sense of "convergence in probability".

Recall the convergence of a deterministic sequence.

Let $a_1, a_2, \ldots$ be a sequence of real numbers and let a be another real. We say that a_n converges to a or $\lim_{n\to\infty} a_n = a$ if for every $\epsilon > 0$ there exists some n_0 s.t.

$|a_n - a| \leq \varepsilon$ for all $n \geq n_0$.



Convergence in Probability

Let $Y_1, Y_2, \ldots$ be a sequence of random variables, and let a be a real number. We say that Y_n converges to a in probability if for every $\varepsilon > 0$ we have

$$\lim_{n \to \infty} P(|Y_n - a| \geq \varepsilon) = 0.$$

- If RVs $Y_1, Y_2, \ldots$ have a PMF or a PDF and converge in probability to a, then "almost all" of the PMF or PDF of Y_n is concentrated within ε of a for large values of n.
- The above condition of convergence in probability can be equivalently written as follows: for every $\varepsilon > 0$ and for every $\delta > 0$ there exists some n_0 such that

$$P(|Y_n - a| \geq \varepsilon) \leq \delta \quad \text{for all } n \geq n_0.$$

$\varepsilon \to$ accuracy level, $\delta \to$ confidence level.

Example. Let $X_1, X_2, \ldots$ be a sequence of independent and uniformly distributed RVs in $[0,1]$, and let $Y_n = \min\{X_1, \ldots, X_n\}$.

Note that $Y_1 \geq Y_2 \geq \ldots$. In fact,

$$Y_n \to 0 \quad \text{in probability.}$$

$$P(|Y_n - 0| \geq \varepsilon) = P(X_1 \geq \varepsilon, \ldots, X_n \geq \varepsilon)$$

$$= (1-\varepsilon)^n \to 0 \quad \text{as } n \to \infty$$

for every $\varepsilon > 0$.

- The following example shows that if $Y_n \to a$, then $E[Y_n]$ may not converge to a.

Example. $$p(Y_n = y) = \begin{cases} 1 - 1/n, & \text{for } y = 0 \\ 1/n, & \text{for } y = n^2 \\ 0, & \text{elsewhere.} \end{cases}$$

For every $\varepsilon > 0$ $\quad P(|Y_n - 0| \geq \varepsilon) = \frac{1}{n} \to 0$ as $n \to \infty$.

$E[Y_n] = n \to \infty$ as $n \to \infty$,

i.e. $E[Y_n] \not\to a$.

Central Limit Theorem (CLT)

Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed RVs with common mean μ and variance σ^2.

Define $Z_n = \dfrac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma}$. Then

$$\lim_{n\to\infty} P(Z_n \le z) = \Phi(z)$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx.$$

- $S_n = \sum_{i=1}^{n} X_i$, $\mathrm{Var}(S_n) = n\sigma^2 \to \infty$ as $n \to \infty$

$M_n = \sum_{i=1}^{n} X_i / n$, $\mathrm{Var}(M_n) = \dfrac{\sigma^2}{n} \to 0$ as $n \to \infty$

$Z_n = \sum_{i=1}^{n} X_i / \sqrt{n}$, $\mathrm{Var}(Z_n) = 1$ independent of n (constant)

Proof of CLT:

Fact. If $M_{Z_n}(t) \to M_Z(t)$, then

$$F_{Z_n}(z) \to F_Z(z) \text{ for all } z.$$

Proof of the fact is related to the continuity of inverse Fourier transform.

We assume that $M_X(s)$ is finite when $-d < s < d$. Let X_i has zero mean & variance 1.

$$Z_n = \sum_{i=1}^{n} X_i / \sqrt{n}.$$

$$M_{Z_n}(s) = E\left[e^{s \sum_{i=1}^{n} X_i/\sqrt{n}} \right]$$

$$= \prod_{i=1}^{n} E\left[e^{s X_i/\sqrt{n}} \right] = M_X(s/\sqrt{n})^n.$$

Let $L(t) = \log M_X(t)$.

$\mathcal{L}(0) = 0, \quad \mathcal{L}'(0) = \dfrac{M_x'(0)}{M_x(0)} = E[X] = 0.$

$\mathcal{L}''(0) = \dfrac{M_x(0)M_x''(0) - M_x'(0)^2}{M_x(0)^2} = E[X^2] = 1.$

We first show that

$$M_x(s/\sqrt{n})^n \longrightarrow e^{t^2/2},$$

↑ MGF of $N(0,1)$

or equivalently, that

$$n\mathcal{L}(s/\sqrt{n}) \longrightarrow t^2/2 \quad \text{as} \quad n \to \infty.$$

$$\lim_{n\to\infty} \frac{\mathcal{L}(s/\sqrt{n})}{n^{-1}} = \lim_{n\to\infty} \frac{-\mathcal{L}'(s/\sqrt{n}) \cdot s\, n^{-3/2}}{-2\, n^{-2}}$$

(by L'Hopital Rule)

$$= \lim_{n\to\infty} \frac{\mathcal{L}'(s/\sqrt{n}) \cdot s}{2n^{-1/2}}$$

$$= \lim_{n\to\infty} \frac{-\mathcal{L}''(s/\sqrt{n})\, s^2\, n^{-3/2}}{-2n^{-3/2}} = \lim_{n\to\infty} \mathcal{L}''(s/\sqrt{n}) \frac{s^2}{2} = \frac{s^2}{2},$$

(by L'Hopital rule) ($\because \mathcal{L}''(0) = 1$)

If x_i's are of mean μ & variance σ^2,

$$\frac{\sum_{i=1}^{n} x_i - n\mu}{\sqrt{n}\sigma} = \frac{\sum_{i=1}^{n} \frac{x_i - \mu}{\sigma}}{\sqrt{n}} \longrightarrow \begin{matrix} \text{mean } 0 \\ \text{var} = 1. \end{matrix}$$

Fact. If $M_{Z_n}(t) \to M_Z(t)$, then

$$F_{Z_n}(z) \to F_Z(z) \text{ for all } z.$$

Proof is related inverse Fourier transform and its continuity.

This completes the proof of the central limit theorem.

Lecture 23

(9 November 2023)

Central Limit Theorem.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables with common mean μ and variance σ^2. Define

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}.$$ Then the CDF of Z_n converges to the standard normal CDF in the sense that

$$\lim_{n\to\infty} P(Z_n \le z) = \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx,$$

for every z.

Proof. We prove this for $\mu = 0$ & $\sigma^2 = 1$, i.e., $Z_n = \sum_{i=1}^{n} X_i/\sqrt{n}$.

$$M_{Z_n}(s) = E\left[e^{sZ_n}\right]$$

$$= E\left[e^{s\sum_{i=1}^{n} x_i/\sqrt{n}}\right]$$

$$= \prod_{i=1}^{n} M_{x_i}(s/\sqrt{n}) = \left(M_x(s/\sqrt{n})\right)^n.$$

Let $L(s) = \log M_x(s)$.

$$\log M_{Z_n}(s) = n \log M_x(s/\sqrt{n}) = n L(s/\sqrt{n}).$$

We have $L(0) = L'(0) = 0$ & $L''(0) = 1$.

Consider $\lim_{n\to\infty} n \log M_x(s/\sqrt{n})$

$$= \lim_{n\to\infty} \frac{\log M_x(s/\sqrt{n})}{n^{-1}}$$

$$= \lim_{n\to\infty} \frac{M_x'(s/\sqrt{n})\, s\, n^{-3/2}}{2 M_x(s/\sqrt{n})\, n^{-2}}$$

$$= \lim_{n\to\infty} \frac{L'(s/\sqrt{n})\, s}{2\, n^{-1/2}}$$

$$= \lim_{n\to\infty} \frac{L''(s/\sqrt{n})\, s^2\, n^{-3/2}}{2\, n^{-3/2}} = s^2/2.$$

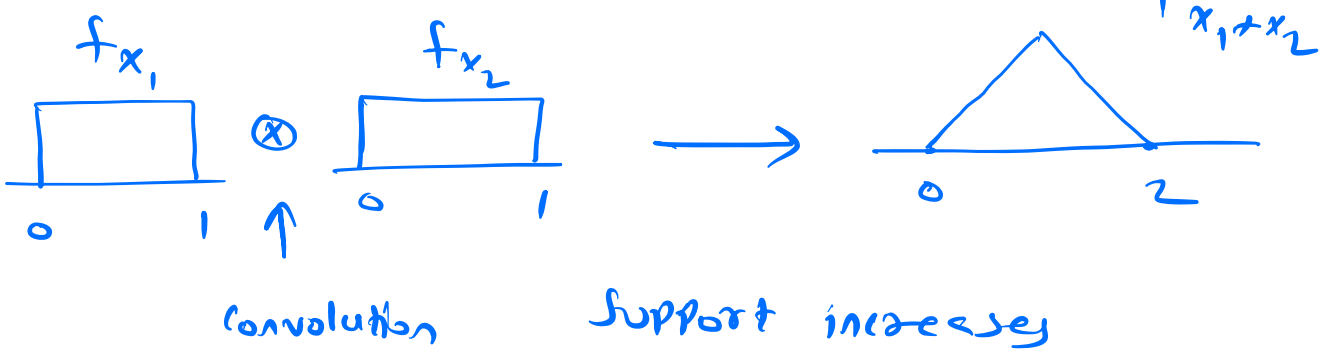
Examples. (i) $X_i \sim N(\mu, \sigma^2)$

$$\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \sim N(0,1) \quad \text{for all } n \in \mathbb{N},$$

$$\frac{X_1 - \mu}{\sigma}, \quad \frac{X_1 + X_2 - 2\mu}{\sqrt{2}\,\sigma}, \quad \cdots \quad N(0,1),$$

$$\lim_{n \to \infty} P(Z_n \le z) = P(Z \le z), \quad Z \sim N(0,1).$$

(ii) $X_i \sim$ uniform $[0,1]$.



convolution support increases

Intuitively as $n \to \infty$, support will be real line and $\sum_{i=1}^{n} X_i \to$ Gaussian RV.

Normal Approximation Based on CLT:

Let $S_n = \sum_{i=1}^{n} x_i$, where x_i are i.i.d. with mean μ and variance σ^2. We are interested in computing $P(S_n \leq c)$. If n is large it can be approximated in the following way.

$$P(S_n \leq c) = P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \leq \frac{c - n\mu}{\sqrt{n}\sigma}\right)$$

$$\approx \Phi(z) \quad \text{where } z = \frac{c - n\mu}{\sqrt{n}\sigma}.$$

Example. X_i, $i \in [1:100]$ are i.i.d. and uniformly distributed in the interval $[5, 50]$. Find $P(S_{100} > 3000)$, $S_{100} = \sum_{i=1}^{100} x_i$. It is not easy to calculate CDF and the desired probability exactly, but an approximate answer can be quickly obtained using CLT.

$$\mu = \frac{5+50}{2}, \quad \sigma^2 = \frac{(50-5)^2}{12} = 168.75.$$

$$P(S_{100} > 3000) = 1 - P(S_{100} \leq 3000)$$

$$= 1 - P\left(\frac{S_{100} - n\mu}{\sqrt{n}\sigma} \leq \frac{3000 - 275}{10(168.75)}\right)$$

$$= 1 - P\left(\frac{S_{100} - n\mu}{\sqrt{n}\sigma} \leq 1.92\right)$$

$$= 1 - \Phi(1.92)$$

$$\approx 1 - 0.9726$$

(using standard Normal table)

$$= 0.0274.$$

Example (Polling). $X_i = \begin{cases} 1 & \text{if } i^{th} \text{ voter votes for A} \\ 0 & \text{otherwise} \end{cases}$

$E[X_i] = p$ $\quad M_n = \sum_{i=1}^{n} X_i / n$.

WLLN (or Chebyshev's inequality):

For $n = 100$, we got that

$$P(|M_n - p| \geq 0.1) \leq 0.25.$$

We will see now if CLT can improve this bound.

$$P\left(\left|\frac{\sum_{i=1}^{n} x_i}{n} - p\right| \geq 0.1\right)$$

$$= P\left(\left|\frac{\sum_{i=1}^{n} x_i - np}{n}\right| \geq 0.1n\right)$$

$$= P\left(\left|\frac{\sum_{i=1}^{n} x_i - np}{\sqrt{n}\sigma}\right| \geq \frac{0.1\sqrt{n}}{\sigma}\right)$$

$$= P\left(\frac{\sum_{i=1}^{n} x_i - np}{\sqrt{n}\sigma} \geq \frac{0.1\sqrt{n}}{\sigma}\right) + P\left(\frac{\sum_{i=1}^{n} x_i - np}{\sqrt{n}\sigma} \leq \frac{-0.1\sqrt{n}}{\sigma}\right)$$

$$\approx 2P\left(Z \geq \frac{0.1\sqrt{n}}{\sigma}\right) \quad Z \sim N(0,1)$$

$$\leq 2P\left(Z \geq 0.2\sqrt{n}\right) \quad \text{as } \sigma \leq \frac{1}{2}$$

$$= 2 - 2P\left(Z \leq 0.2\sqrt{n}\right)$$

$$= 2 - 2\Phi\left(0.2\sqrt{100}\right) = 2 - 2\Phi(2) = 2 - 2 \times 0.977$$

$$= 0.046.$$

This is a much better estimate than 0.25 from WLLN.

Strong Law of Large Numbers

- SLLN also deals with the convergence of the sample mean to the true mean. However, SLLN refers to a different type of convergence.

WLLN

$$M_n = \sum_{i=1}^{n} x_i / n \longrightarrow E[x] \quad \text{in Probability}$$

SLLN

$$M_n = \sum_{i=1}^{n} x_i / n \longrightarrow E[x] \text{ with probability 1.}$$

(or almost sure convergence)

Almost sure convergence:

Let $Y_1, Y_2, \ldots$ be a sequence of RVs and c be a real number. Then we say $Y_n \xrightarrow{a.s.} a$ if $P\left(\left\{\omega \in \Omega : \lim_{n \to \infty} x_n(\omega) = c\right\}\right) = 1$.

Example. Let $\Omega = [0,1]$. Consider a probability law defined by $P([a,b]) = b-a$ for all $0 \le a \le b \le 1$.

Define $X_n(\omega) = \omega^n$ for $n \in \mathbb{N}$.

Note that $\lim_{n\to\infty} X_n(\omega) = \begin{cases} 0, & \text{if } 0 \le \omega < 1 \\ 1, & \text{if } \omega = 1. \end{cases}$

always find this first

$$P(\{\omega : \lim_{n\to\infty} X_n(\omega) = 0\}) = P([0,1)) = 1$$

Since the singleton set $\{1\}$ has zero probability,

$$\therefore X_n \xrightarrow{a.s.} 0.$$

SLLN. Let $X_1, X_2, \ldots$ be a sequence of i.i.d. RVs with mean μ. Then,

$$M_n = \frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow{a.s.} \mu, \quad \text{i.e.,}$$

$$P\left(\{\omega : \lim_{n\to\infty} M_n(\omega) = \mu\}\right) = 1.$$

Proof. Assume that $E[X_i^4] = k < \infty$,

$S_n = \sum_{i=1}^{n} X_i$. $E[S_n^4] = E\left[\left(\sum_{i=1}^{n} X_i\right)^4\right]$.

This will have terms of the form

$$X_i^4, \quad X_i^3 X_j, \quad X_i^2 X_j^2, \quad X_i^2 X_j X_k, \quad X_i X_j X_k X_l$$

where i, j, k, l are different.

Assume $\mu = 0$. Then because of independence it follows that

$$E[X_i^3 X_j] = E[X_i^3] E[X_j] = 0$$

$$E[X_i^2 X_j X_k] = E[X_i^2] E[X_j] E[X_k] = 0$$

$$E[X_i X_j X_k X_l] = E[X_i] E[X_j] E[X_k] E[X_l] = 0.$$

so $E[S_n^4] = n E[X_i^4] + 6 \binom{n}{2} E[X_i^2 X_j^2]$

$$= n E[X_i^4] + 3n(n-1) \underbrace{E[X_i^2] E[X_j^2]}$$

$$= (E[X_i^2])^2 \le E[X_i^4]$$

as $\mathrm{Var}(X_i^2) \ge 0$

$$\le nk + 3n(n-1)k \le 3n^2 k$$

$$\Rightarrow E[S_n^4 / n^4] \le 3k / n^2$$

$$\Rightarrow E\left[\sum_{n=1}^{\infty} S_n^4/n^4\right] = \sum_{n=1}^{\infty} E\left[S_n^4/n^4\right]$$

$$= \sum_{n=1}^{\infty} \frac{1}{n^2} \cdot (3K) < \infty.$$

This implies that

$$P\left(\sum_{n=1}^{\infty} S_n^4/n^4 < \infty\right) = 1$$

$$\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} < \infty \Rightarrow \lim_{n\to\infty} \frac{S_n^4}{n^4} = 0$$

$$\Rightarrow 1 = P\left(\sum_{n=1}^{\infty} \frac{S_n^4}{n^4} < \infty\right) \leq P\left(\lim_{n\to\infty} \frac{S_n^4}{n^4} = 0\right)$$

$$\Rightarrow P\left(\lim_{n\to\infty} \frac{S_n}{n} = 0\right) = 1.$$

In the above proof we have used

$E\left[\sum_{i=1}^{\infty} Z_i\right] = \sum_{i=1}^{\infty} E[Z_i]$. This is not necessarily true always for any Z_i's.

However this holds true when all Z_i are non-negative random variables. This is because of monotone convergence theorem (not covered in this course).

Lecture 24

(16 November 2023)

Random Processes

- A random or stochastic process is a mathematical model of a probabilistic experiment that evolves in time and generates a sequence of numerical values.
- Each numerical value in the sequence is modeled by a random variable, so a random process is simply a (finite or infinite) sequence of random variables.

Recall that a RV $X : \Omega \to \mathbb{R}$.

Formally, a random process is a family of random variables $(X_t : t \in T)$, all on the same probabilistic model (Ω, P).

In many applications, T is a set of times. If $T = \mathbb{N}$, then $(X_t : t \in T)$ is called a discrete-time process. If $T = \mathbb{R}$, then continuous-time process.

- For each $t \in T$ X_t is a RV.

$\omega \mapsto X_1(\omega)\ X_2(\omega)\ \ldots$

Discrete-time

$\omega \mapsto X_t(\omega)\ t \in T.$

Continuous-time

For a fixed $\omega \in \Omega$ $(X_t(\omega)\ t \in T)$ is called the sample path at ω.

Mean Function:

$$\mu_X(t) = E[X_t] \text{ for } t \in T.$$

Covariance Function:

$$C_X(t_1, t_2) = \text{Cov}(X_{t_1}, X_{t_2})$$

$$= E[X_{t_1} X_{t_2}] - E[X_{t_1}] E[X_{t_2}]$$

Correlation function:

$$R_X(t_1, t_2) = E[X_{t_1} X_{t_2}].$$

- A random process is statistically specified by its complete set of n^{th} order distribution function for all $n \in N$.

$$F_x(x_1 \dots x_n ; t_1 t_2 \dots t_n) = P(X_{t_1} \le x_1, \dots, X_{t_n} \le x_n).$$

For the continuous case we can obtain the PDF as

$$f(x_1 x_2 ; t_1 t_2) = \frac{\partial^2 F(x_1 x_2 ; t_1 t_2)}{\partial x_1 \partial x_2}.$$

Arrival-Type Processes we are interested in occurances that have the character of an arrival such as message receptions at a receiver, customer purchases at a store etc. we will focus on models in which the interarrival times (the times between successive arrivals) are independent random variables

Bernoulli process - case where arrivals occur in discrete time and the interarrival times are geometrically distributed

Poisson process - case where arrivals occur in continuous time and the interarrival times are exponentially distributed.

The Bernoulli process

Consider a sequence of independent coin tosses with the probability of heads p in the range $0 < p < 1$.

- A Bernoulli process is a sequence $X_1, X_2, \ldots$ of independent Bernoulli RVs X_i with

$$P(X_i = 1) = P(\text{success at the } i\text{th trial})$$

$$P(X_i = 0) = P(\text{failure at the } i\text{th trail})$$

- No. of arrivals within a certain time period or no. of successes in n independent trails:

$$P_S(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad k = 0, 1, \ldots, n$$

$$E[S] = np, \quad \text{var}(S) = np(1-p).$$

This is Binomial with parameters n & p.

- The time until the first arrival or no. of trials up to (and including) the first success:

$$p_T(t) = (1-p)^{t-1} p, \quad t = 1, 2, \ldots$$

$$E[T] = \frac{1}{p}, \quad \text{var}(T) = \frac{1-p}{p^2}.$$

Independence & Memorylessness

Due to independence property in the Bernoulli process whatever has happened in the past trials provides no information on the outcome of future trials.

Fresh-start property:

For any given n, consider

$X_1, X_2, \ldots, X_n, X_{n+1}, X_{n+2}, \ldots$, we notice that

$Y_i = X_{n+i}$ are independent Bernoulli trials and therefore form a Bernoulli process.

Memorylessness property: T = time until 1st success

$$P(T-n=t \mid T>n)$$

$$= \frac{P(T-n=t \text{ and } T>n)}{P(T>n)}$$

$$= \frac{P(T=n+t)}{P(T>n)} = \frac{(1-p)^{n+t-1} \cdot p}{(1-p)^n}$$

$$= (1-p)^{t-1} p = P(T=t).$$

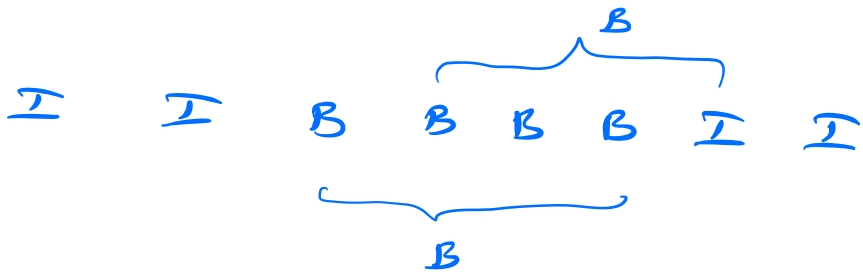
- we call a slot busy (B) if there is an arrival, and otherwise let us call it idle. we call a string of idle slots flanked by busy slots an 'idle period'. similarly we can define 'busy period'.

Let us derive the PMF, mean, and variance of the following random variables.

T = the time index of the first idle slot

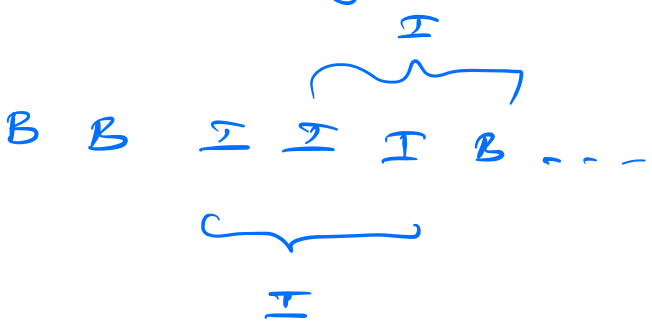
$$P_T(k) = p^{k-1}(1-p) \quad k = 1, 2, \ldots$$

L = the length (no. of slots) of the first busy period



$$P_B(k) = p^{k-1}(1-p) \quad k = 1, 2, \ldots$$

I = the length of the first idle period



$$P_I(k) = (1-p)^{k-1} p \quad k = 1, 2, \ldots$$

Example. Let N be the first time that we have a success immediately following a previous success. That is, N is the first time i for which $x_{i-1} = x_i = 1$. What is the probability that there are no successes in the two trials that follow, i.e., $P(x_{N+1} = x_{N+2} = 0)$?

$$P(X_{N+1} = X_{N+2} = 0)$$

$$= \sum_{n=1}^{\infty} P(X_{N+1} = X_{N+2} = 0 \mid N = n)\, P(N = n)$$

$$P(X_{N+1} = X_{N+2} = 0 \mid N = n)$$

$$= P(X_{n+1} = X_{n+2} = 0) = (1-p)^2$$

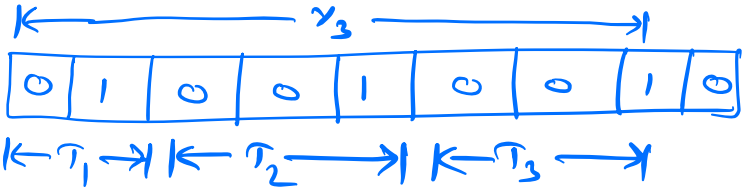
(equality because $N = n$ is determined by some function of $X_1, X_2, \ldots, X_n$).

Interarrival Times

Let Y_k denote the time of the k^{th} success (or arrival).

$$T_1 = Y_1 \qquad T_k = Y_k - Y_{k-1}, \quad k = 2, 3, \ldots$$

T_k represents the k^{th} interarrival time i.e. no. of trials following the $(k-1)^{st}$ success until the next success.



We have $Y_k = \sum_{i=1}^{k} T_i$.

Interarrival times $T_1, T_2, T_3, \ldots$ are independent and all have the same geometric distribution.

Properties of the k^{th} Arrival Time

$$Y_k = T_1 + T_2 + \cdots + T_k$$

Each T_i is a geometric RV with parameter p and all T_i's are independent.

$$E[Y_k] = \sum_{i=1}^{k} E[T_i] = k/p,$$

$$\text{Var}(Y_k) = \sum_{i=1}^{k} \text{Var}(T_i) = \frac{k(1-p)}{p^2},$$

To find the pmf of Y_k

$$P(Y_k = t) = P(A \cap B) \quad \text{where}$$

$A = \{$ trial t is a success $\}$

$B = \{$ exactly $k-1$ successes in $t-1$ trials $\}$

$$P_{Y_k}(t) = P(A) \cdot P(B)$$

$$= p \cdot \binom{t-1}{k-1} p^{k-1} (1-p)^{t-k}$$

$$= \binom{t-1}{k-1} p^{k} (1-p)^{t-k}.$$

Known as Pascal PMF of order k.

Poisson Process

- A counting process counts the number of arrivals, e.g., no. of customers who arrive at a restaurant — $X(t)$ = no. of arrivals in time t. $[X(0) = 0]$
- A counting process has independent and stationary increments.

Independent increments:

For all $0 \le t_1 < t_2 < t_3 \cdots < t_n$ the RVs

$X(t_2) - X(t_1)$, $X(t_3) - X(t_2)$, ... are independent.

Note that $X(t_i) - X(t_{i-1})$ represents the no. of arrivals in the interval $(t_{i-1}, t_i]$.

Stationary increments:

For all $t_2 > t_1 \geq 0$ & $\tau > 0$, the RVs $X(t_2) - X(t_1)$ and $X(t_2+\tau) - X(t_1+\tau)$ have the same distribution.

- Thus a counting process has independent increments if the no. of arrivals in non-overlapping intervals are independent.

Also it has stationary increments if, for all $t_2 > t_1 \geq 0$ $X(t_2) - X(t_1)$ has the same distribution as $X(t_2 - t_1)$.

- Before we define Poisson process we recall Poisson RV.

$X \sim \text{Poisson}(\lambda)$ means $P_X(k) = \frac{e^{-\lambda} \lambda^k}{k!}$, $k = 0, 1, 2, \ldots$

- A counting process $(N(t), t \in [0, \infty))$ is called a Poisson process with rate λ if:

(1) $N(0) = 0$.

(2) $N(t)$ has independent increments.

(3) the no. of arrivals in any interval of length $T > 0$ has Poisson (λT) distribution.

$E[N(t)] = \lambda t$ $\quad$ $\text{var}(N(t)) = \lambda t$.

$$R_X(t_1, t_2) = E[N(t_1) N(t_2)] \quad (t_1 < t_2)$$

$$= E[N(t_1)(N(t_2) - N(t_1) + N(t_1))]$$

$$= E[N(t_1)(N(t_2) - N(t_1))] + E[N^2(t_1)]$$

$$= E[N(t_1)] E[N(t_2 - t_1)] + E[N^2(t_1)]$$

$$= \lambda t_1 \cdot \lambda(t_2 - t_1) + \lambda t_1 + \lambda^2 t_1^2$$

$$= \lambda^2 t_1 t_2 - \lambda^2 t_1^2 + \lambda t_1 - \lambda^2 t_1^2$$

$$= \lambda t_1 + \lambda^2 t_1 t_2.$$

$$C(t_1, t_2) = \lambda \min\{t_1, t_2\}.$$

Lecture 25

(18 November 2023)

- A random process $X = (X_t : t \in T)$ is specified by n^{th} order distribution function

$$F_{X_{t_1} \cdots X_{t_n}}(x_1, x_2, \cdots, x_n)$$

$$= P(X_{t_1} \le x_1, \cdots, X_{t_n} \le x_n)$$

for all $n \in \mathbb{N}$.

Example. Consider the random process

$$X(t) = A \sin(\omega_0 t + \Theta),$$

where A and Θ are independent and Θ is uniformly distributed over $[-\pi, \pi]$.

$$\mu_X(t) = E[A \sin(\omega_0 t + \Theta)]$$

$$= \mu_A \cdot E[\sin(\omega_0 t + \Theta)]$$

$$= \mu_A \cdot \int_{-\pi}^{\pi} \frac{1}{2\pi} \cdot \sin(\omega_0 t + \theta)\, d\theta$$

$= 0.$

$$R_X(t_1, t_2) = E[X(t_1) X(t_2)]$$

$$= E[A^2 \sin(\omega_0 t_1 + \Theta) \sin(\omega_0 t_2 + \Theta)]$$

$$= E[A^2] \; E[\sin(\omega_0 t_1 + \Theta) \sin(\omega_0 t_2 + \Theta)]$$

$$= E[A^2] \frac{1}{2} \left(E[\cos(\omega_0 (t_1 - t_2))] - \underbrace{E[\cos(\omega_0 (t_1 + t_2) + 2\Theta)]}_{= 0} \right)$$

$$= \frac{1}{2} E[A^2] \; [\cos(\omega_0 (t_1 - t_2))].$$

Stationary Process

A random process $(X(t), t \in R)$ is stationary if for all $t_1, t_2, \ldots, t_n, T \in R$

$$F_{X(t_1) X(t_2) \ldots X(t_n)}(x_1, x_2, \ldots, x_n) = F_{X(t_1+T) \ldots X(t_n+T)}(x_1, x_2, \ldots, x_n)$$

$\forall n \in N.$

Exercise. Write down a similar definition for discrete-time random process

Example. Consider i.i.d. discrete-time random process $X_1, X_2, \ldots$. Is this stationary?

Yes!

$(X_{n_1}, X_{n_2}, \ldots, X_{n_k})$ has same distribution as

$(X_{n_1+\tau}, \ldots, X_{n_k+\tau})$.

— If a process is stationary the analysis is usually simpler as the probabilistic properties do not change by time.

However it turns out that not many real-life processes are stationary. Even if a process is stationary it might be difficult to prove it. Fortunately often a weaker notion of stationarity suffices.

Wide-Sense stationary (WSS)

A random process $X(t)$ is called WSS if for all $t, t_1, t_2 \in R$, $T \in R$,

$$\mu_X(t) = \mu_X(t+T)$$

$$R_X(t_1, t_2) = R_X(t_1+T, t_2+T)$$

so $\mu_X(t)$ is a constant and $R_X(t_1, t_2)$ is a function of $t_2 - t_1$.

Example. $X(t) = A \sin(\omega_0 t + \theta)$ A, θ are independent & $\theta \sim$ uniform $[-\pi, \pi]$.

$X(t)$ is WSS because

$$\mu_X(t) = 0 \quad (\text{a constant})$$

$$R_X(t_1, t_2) = \tfrac{1}{2} E[A^2] \cos(\omega_0 (t_1 - t_2))$$

$$= R_X(t_1 - t_2).$$

- For WSS random process it suffices to denote autocorrelation by $R_X(\tau)$.

Exercise. Prove that a stationary process is WSS.

Properties of $R_x(\tau)$

- $R_x(0) = E[x(t)\,x(t)] = E[x^2(t)] \geq 0$.

  $E[x^2(t)]$ is called expected (or average) power in $x(t)$ at time t. For WSS this is not a function of time.

- $R_x(-\tau) = E[x(t+\tau)\,x(t)]$

  $= E[x(t)\,x(t+\tau)]$

  $= R_x(\tau)$.

- $|R_x(\tau)| \leq R_x(0)$ for all $\tau \in R$.

  This follows from Cauchy-Schwarz's inequality,

  $$|E[xy]| \leq \sqrt{E[x^2]\,E[y^2]}$$

  Consider $E[(x-\alpha y)^2] \geq 0$

  $\Rightarrow E[x^2 + \alpha^2 y^2 - 2\alpha xy] \geq 0$

  $\Rightarrow \alpha^2 E[y^2] - 2\alpha E[xy] + E[x^2] \geq 0$

  So discriminant ≤ 0

  $\Rightarrow 4E^2[xy] \leq 4E[x^2]E[y^2]$

$$\Rightarrow |E[XY]| \le \sqrt{E[X^2]E[Y^2]}.$$

$$|E[x(t)x(t+\tau)]| \le \sqrt{E[x^2(t)]\,E[x^2(t+\tau)]}$$

$$\Rightarrow |R_x(\tau)| \le \sqrt{R_x(0)\cdot R_x(0)} = R_x(0).$$

Power Spectral Density (PSD)

Power spectral density is the Fourier transform of $R_x(\tau)$, mathematically

$$S_x(f) = \int_{-\infty}^{\infty} R_x(\tau)\, e^{-j2\pi f\tau}\, d\tau, \quad j = \sqrt{-1}$$

is PSD of $x(t)$.

It can be shown that the inverse Fourier transform of $S(f)$ is $R_x(\tau)$.

$$\int_{-\infty}^{\infty} S_x(f)\, e^{j2\pi f\tau}\, df = R_x(\tau).$$

Properties of PSD

- $S_x(-f) = S_x(f)$.

  This is because $R_x(\tau)$ is even.

- $E[x(t)^2] = R_x(0) = \int_{-\infty}^{\infty} S_x(f)\,df$.

  ↑ Inverse Fourier Transform

  so we get expected or average power in $x(t)$ by integrating the PSD of $x(t)$. This is why $S_x(f)$ is called power spectral density.

- $S_x(f) \geq 0$.

  (Proof omitted)

Lecture 26

(20 November 2023)

Q) (a) Is an event A independent of itself? No in general.

$P(A \cap A) = P(A)^2 \Rightarrow P(A) = 0$ or 1.

(b) Find the probability that exactly one of the events A or B occurs in terms of $P(A)$ $P(B)$ $P(A \cap B)$.

$$P\left((A \cap B^c) \cup (A^c \cap B)\right) = P(A \cap B^c) + P(A^c \cap B)$$

$$= P(A) + P(B) - 2P(A \cap B)$$

(c) Is it true that $P(A \mid A \cup B) \geq P(A \mid B)$?

$$P(A \mid A \cup B) = P(A \mid (A \cup B) \cap B)\, P(B \mid A \cup B) + P(A \mid (A \cup B) \cap B^c)\, P(B^c \mid A \cup B)$$

$$= P(A \mid B)\, P(B \mid A \cup B) + \underbrace{P(A \mid A \cap B^c)}_{=1} P(B^c \mid A \cup B)$$

$$= \lambda P(A|B) + (1-\lambda) 1$$

$$P(A|A\cup B)$$

$$P(A|B) \qquad 1$$

$$\therefore P(A|A\cup B) \geq P(A|B).$$

Q) If X is a positive integer valued RV that satisfies memorylessness property

$$P(X > m+n \mid X > m) = P(X > n), \text{ for any } m, n \in \mathbb{N}.$$

Then prove that X is a geometric random variable.

Proof. Let $P(X > n) = a_n$ for $n \in \mathbb{N}$.

The memorylessness property gives

$$\frac{a_{m+n}}{a_m} = a_n$$

$$\Rightarrow a_{m+n} = a_m a_n \quad \forall\, m, n \in \mathbb{N}$$

$$\Rightarrow a_{m+1} = a_m a_1 = a_1^{m+1}$$

where $a_1 = P(X > 1) = 1 - P(X = 1) \triangleq 1 - p.$

$$P(X = n) = a_{n-1} - a_n = (1-p)^{n-1} - (1-p)^n = (1-p)^{n-1} p.$$

Q) Let X be a discrete RV, Y be a continuous RV and I is a binary RV s.t. X is independent of I

Y is independent of I.

Define

$$Z = \begin{cases} X & \text{if } I=1 \\ Y & \text{if } I=0 \end{cases}.$$

$$Z = \begin{cases} 1 & \text{if } I=1 \\ \text{Uniform}[0,2] & \text{if } I=0 \end{cases}.$$

Z is neither a discrete nor a continuous RV.

$$F_Z(z) = P(Z \le z)$$

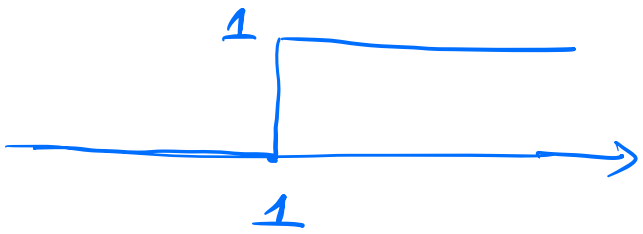
$$= P(Z \le z \mid I=1)P(I=1) + P(Z \le z \mid I=0)P(I=0)$$

$$= p F_X(z) + (1-p) F_Y(z)$$

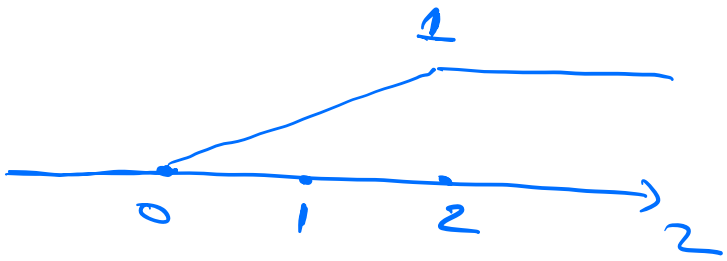
$(p = P(I=1))$

Z is a mixed random variable.

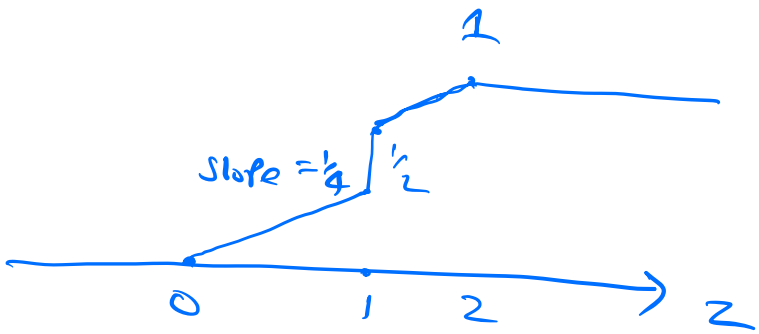
$F_x(z)$



$F_y(z)$



For $p = 1/2$ $f_z(z) =$



Application of Probability in Information Leakage.

The "information content" of a message depends on the degree to which the content of the message is surprising. If a highly likely event occurs, the message carries very little information. If a highly

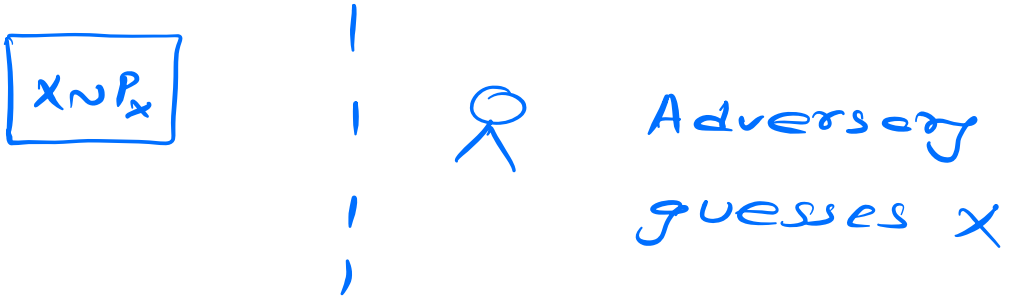
unlikely event occurs the message is more informative.

The information content in an event with probability $p = \log 1/p$.

[all logarithms are to the base 2]

Vulnerability: Suppose $x \sim P_x$. The vulnerability of x is given by

$$V(x) = \max_{x \in \mathcal{X}} P_x(x).$$



$V(x)$ is the worst-case probability that an adversary could guess the value of x correctly.

$\log \frac{1}{V(x)}$ can be viewed as information measure

min-entropy of x $H_\infty(x) = \log \frac{1}{V(x)}$

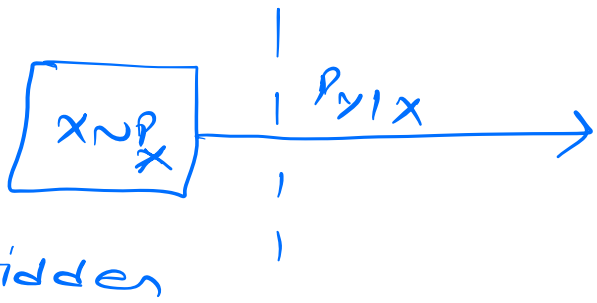
$$= \log \frac{1}{\max_x P_x(x)}.$$

— Given two RVs $(X, Y) \sim P_{XY}$, the conditional vulnerability

$$V(X|Y) = \sum_{y \in \mathcal{Y}} P_Y(y)\, V(X|Y=y)$$

$$= \sum_{y \in \mathcal{Y}} P_Y(y) \max_x P_{X|Y}(x|y).$$

$$H_\infty(X|Y) = \log \frac{1}{V(X|Y)}.$$



information leaked about X?

Information leaked about X from Y

= initial uncertainity in X

—

remaining uncertainity in X after observing Y

Leakage $\mathcal{L}_{P_X}(X \to Y) = H_\infty(X) - H_\infty(X|Y)$

$$= \log \frac{V(X|Y)}{V(X)}.$$

If P_X is unknown the worst-case leakage is of interest.

Theorem. $\max_{P_X} \mathcal{L}_{P_X}(X \to Y) = \mathcal{L}_{P_U}(X \to Y)$

$$= \sum_y \max_x P_{Y|X}(y|x).$$

(Sum of maximum of the columns of $P_{Y|X}$ matrix)

Proof. $2^{\mathcal{L}_{P_X}(X \to Y)} = \frac{V(X|Y)}{V(X)}$

$$= \sum_y P_Y(y) \max_x P_{X|Y}(x|y) \Big/ \max_{x'} P_X(x')$$

$$= \sum_y P_Y(y) \max_x \frac{P_{Y|X}(y|x) P_X(x)}{P_Y(y)} \Big/ \max_{x'} P_X(x')$$

$$= \sum_y \frac{\max_x P_{Y|X}(y|x) P_X(x)}{\max_{x'} P_X(x')}$$

$$\leq \sum_y \max_x P_{Y|X}(y|x) \cdot \frac{\max_{x'} P_X(x')}{\max_{x'} P_X(x')}$$

$$= 2^{L_{P_U}(X \to Y)},$$

where $P_U(x) = \frac{1}{|\mathcal{X}|}, \forall x \in \mathcal{X}$.

Remark. This is an operationally motivated leakage measure and satisfies all the axioms of a leakage measure. Interestingly mutual information $I(X;Y) = H(X) - H(X|Y)$ does not satisfy all.

Q) Let (X,Y) have the joint PDF

$$f_{XY}(x,y) = \begin{cases} x+y, & (x,y) \in [0,1]^2 \\ 0, & \text{otherwise.} \end{cases}$$

Let $Z = X(1+Y)$ $W = X^2$.

$g_1(x,y) = x(1+y)$, $g_2(x,y) = x^2$.

$x = h_1(z,w) = \sqrt{w}$, $y = h_2(z,w) = \frac{z}{\sqrt{w}} - 1$.

$\frac{\partial g_1}{\partial x} = 1+y$, $\frac{\partial g_1}{\partial y} = x$

$\frac{\partial g_2}{\partial x} = 2x$, $\frac{\partial g_2}{\partial y} = 0$.

$$J = \begin{vmatrix} 1+y & x \\ 2x & 0 \end{vmatrix} = 2x^2 = 2w$$

$$f_{zw}(z,w) = \frac{f_{xy}\left(\sqrt{w}, \frac{z}{\sqrt{w}} - 1\right)}{2w}$$

$$= \frac{\sqrt{w} + \frac{z}{\sqrt{w}} - 1}{2w} ,$$

for (z,w) s.t. $0 \le w \le 1$, $\sqrt{w} \le z \le 2\sqrt{w}$.

0 , otherwise.