

STA2603 Distribution Theory II

- Always specify the domain of the function, including elsewhere, always even in the intermediate stages!
- Always specify the parameters when identifying the distribution!

Commonly used series expansions			
	e^{tx}	$= 1 + \frac{tx}{1!} + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \dots$	
$\log(1+x)$	$= \log_e(1+x)$	$= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$	
$\log(1-x)$	$= \log_e(1-x)$	$= -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots$	

Study Unit 2: Random Variables

- Frequency Function (discrete) AKA Mass function
- Density function (continuous).
- Distribution function is cumulative density function

Discrete Random Variables

Distribution	Mass function $p_X(k)$	$M_X(t)$ mgf	Mean & Variance
Bernoulli $ber(p)$ one success	$\begin{cases} p^k (1-p)^{1-k} & k = 0; 1 \\ 0 & \text{elsewhere} \end{cases}$	$(1-p + pe^t)$	$\begin{aligned} \mu_X &= p \\ \sigma_X^2 &= p(1-p) \end{aligned}$
sum of bernoulli Binomial $b(n; p)$ x successes in n trials	$\begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0; 1; \dots; n \\ 0 & \text{elsewhere} \end{cases}$	$(1-p + pe^t)^n$	$\begin{aligned} \mu_X &= np \\ \sigma_X^2 &= np(1-p) \end{aligned}$
Geometric $geo(p)$ 1st success in x trials	$\begin{cases} (1-p)^{x-1} p, & x = 1; 2; 3; \dots \\ 0 & \text{elsewhere} \end{cases}$	$\frac{pe^t}{1 - (1-p)e^t}$	$\begin{aligned} \mu_X &= \frac{1}{p} \\ \sigma_X^2 &= \frac{1-p}{p^2} \end{aligned}$
sum of geometric Negative Binomial $nb(r; p)$ r'th success in k trials	$\begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r}, & k = 1; 2; \dots \\ 0 & \text{elsewhere} \end{cases}$	$\left[\frac{pe^t}{1 - (1-p)e^t} \right]^r$	$\begin{aligned} \mu_X &= \frac{r}{p} \\ \sigma_X^2 &= \frac{r(1-p)}{p^2} \end{aligned}$
limit of binomial Poisson $Po(\lambda)$ K rare events	$\begin{cases} \frac{\lambda^k}{k!} e^{-\lambda}, & k = 0; 1; \dots \\ 0 & \text{elsewhere} \end{cases}$	$e^{\lambda(e^t - 1)}$	$\begin{aligned} \mu_X &= \lambda \\ \sigma_X^2 &= \lambda \end{aligned}$

$$\text{Hypergeometric : } p(x) = \begin{cases} \frac{\binom{r}{x} \binom{n-r}{m-x}}{\binom{n}{m}} & \text{if } x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

Probability of x successes in a sample size m from sample space n and a subset r

1.	$p_X(x_i) \geq 0$
2.	$\sum_{x=0}^{\infty} p_X(x) = 1$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

$$\rightarrow P(X > x) = 1 - P(X \leq x)$$

→ To prove is distribution:

- 1) State probabilities are all non-negative.
- 2) Show probabilities all sum to 1.

$$\lim_{x \rightarrow a^+} F_X(x) = F_X(a) \text{ for all real values of } x$$

Bernoulli Random Variables

$$\text{Indicator random variable } I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \text{otherwise} \end{cases}$$

$$p_X(x) = \begin{cases} 1-p & \text{if } x=0 \\ p & \text{if } x=1 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 < x < 1 \\ 1 & \text{if } x > 1 \end{cases}$$

Binomial Distribution

- The sum of independent Bernoulli variables is a binomial random variable.
- To prove binomial probabilities sums to 1. Use finite binomial series expansion:
- 1) State binomial function sums to 1
 - 2) then given the binomial expansion of $(a+b)^n = \sum_{x=0}^n \binom{n}{x} a^{n-x} b^x$
 - 3) let $a = 1-p$ and $b = p$
 - 4) Substitute above into $(a+b)^n$
 - 5) therefore $(1-p+p)^n = 1^n = 1$

Geometric Distribution

→ To prove that geometric distribution sums to 1, use Taylor series at $x=0$

$$\begin{aligned} (1-p)^{-1} &= 1 + \frac{(-1)}{1!}(-p) + \frac{(-1)(-1-1)}{2!}(-p)^2 + \frac{(-1)(-1-1)(-1-2)}{3!}(-p)^3 + \dots \\ &= 1 + p + p^2 + p^3 + p^4 + \dots \\ p \sum_{x=1}^{\infty} (1-p)^{x-1} &= p \sum_{r=0}^{\infty} (1-p)^r \\ &= p [1 + (1-p) + (1-p)^2 + (1-p)^3 + \dots] \\ &= p (1 - (1-p))^{-1} \\ &= \frac{p}{p} \\ &= 1 \end{aligned}$$

→ To derive CDF

$$F(x) = \sum_{r=0}^x p(1-p)^r = \frac{p(1 - (1-p)^{x+1})}{1 - (1-p)} = 1 - (1-p)^{x+1}$$

Negative Binomial Distribution

A negative binomial random variable can be expressed as the sum of r independent geometric variables.

$$p_X(x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & \text{if } x = r, r+1, r+2, \dots; 0 \leq p \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Hypergeometric distribution

$$p_X(x) = \begin{cases} \frac{\binom{r}{x} \binom{n-r}{m-x}}{\binom{n}{m}} & \text{if } x = 0, 1, 2, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

Poisson Distribution

To show it is a frequency function:

$$\begin{aligned} e^\lambda &= 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \dots \\ &= \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \end{aligned}$$

$$\begin{aligned} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} e^{-\lambda} &= e^\lambda e^{-\lambda} \\ &= 1. \end{aligned}$$

Poisson distribution can be used to approximate binomial distribution, if n is large and p is small.

Continuous Random Variables

1. $F_X(x)$ is a non-decreasing function
2. $\lim_{x \rightarrow -\infty} F_X(x) = 0$
3. $\lim_{x \rightarrow \infty} F_X(x) = 1$
4. $F_X(x)$ is everywhere continuous

$$P(a < X < b) = \int_a^b f_X(x) dx$$

$$\begin{array}{l}
1. \quad f_X(x) \geq 0 \quad \text{and} \\
2. \quad \int_{-\infty}^{\infty} f_X(x) dx = 1
\end{array}$$

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

Distribution	Density function $f_X(x)$	$M_X(t)$ mgf	Mean & Variance
Uniform	$\begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{elsewhere} \end{cases}$	$\frac{(e^{tb} - e^{ta})}{(b-a)t}$	$\begin{aligned} \mu_X &= \frac{(a+b)}{2} \\ \sigma_X^2 &= \frac{(b-a)^2}{12} \end{aligned}$
Gamma $g(\alpha; \lambda)$	$\begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$	$\left(1 - \frac{t}{\lambda}\right)^{-\alpha}, \quad t < \lambda$	$\begin{aligned} \mu_X &= \frac{\alpha}{\lambda} \\ \sigma_X^2 &= \frac{\alpha}{\lambda^2} \end{aligned}$
Exponential (special gamma with $\alpha = 1$)	$\begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$	$\left(1 - \frac{t}{\lambda}\right)^{-1}$	$\begin{aligned} \mu_X &= \frac{1}{\lambda} \\ \sigma_X^2 &= \frac{1}{\lambda^2} \end{aligned}$
Chi-squared $\chi^2(r)$ (special gamma with $\alpha = \frac{n}{2}; \lambda = \frac{1}{2}$)	$\begin{cases} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \geq 0 \\ 0 & \text{elsewhere} \end{cases}$	$(1 - 2t)^{-\frac{n}{2}}$	$\begin{aligned} \mu_X &= n \\ \sigma_X^2 &= 2n \end{aligned}$
Normal $N(\mu; \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$	$\begin{aligned} \mu_X &= \mu \\ \sigma_X^2 &= \sigma^2 \end{aligned}$
Standard normal $N(0; 1)$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \quad -\infty < x < \infty$	$e^{\frac{1}{2}t^2}$	$\begin{aligned} \mu_X &= 0 \\ \sigma_X^2 &= 1 \end{aligned}$
Beta type 1	$\begin{cases} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1-x)^{b-1}, & 0 \leq x \leq 1 \\ 0 & \text{elsewhere} \end{cases}$	Beyond the scope of this module	$\begin{aligned} \mu_X &= \frac{a}{a+b} \\ \sigma_X^2 &= \frac{ab}{(a+b)^2(a+b+1)} \end{aligned}$

Description	Density function $[f_X(x)]$
t-distribution t_n	$\frac{\Gamma[(n+1)/2]}{\Gamma(n/2)\sqrt{n\pi}} \left(1 + \frac{t^2}{n}\right)^{-\frac{1}{2}(n+1)}, \quad -\infty < t < \infty$
F-distribution $F_{m,n}$	$\frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} x^{m/2-1} \left(1 + \frac{m}{n}x\right)^{-(m+n)/2} \quad x \geq 0$
k -th order statistic	$\frac{n!}{(k-1)!(n-k)!} [F_X(x)]^{k-1} [1 - F_X(x)]^{n-k} f_X(x), \quad -\infty < x < \infty$

→ The p th quantile: $F(x_p) = p$ or $P(X \leq x_p) = p$ and $x_p = F^{-1}(p)$

→ $\ln(1) = 0$

$$\rightarrow \int_0^\infty e^{-y} dy = 1$$

Exponential Density

Memoryless proof

$$\begin{aligned}
 P(T > t + s | T > s) &= \frac{P(T > t + s \text{ and } T > s)}{P(T > s)} \\
 &= \frac{P(T > t + s)}{P(T > s)} \\
 &= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} \\
 &= e^{-\lambda t}
 \end{aligned}$$

Gamma Density

Gamma function

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} \quad \text{Properties } \Gamma(\alpha + 1) = \alpha \Gamma(\alpha) \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

$$c^n \Gamma(n) = \int_0^{\infty} x^{n-1} e^{-\frac{x}{c}}$$

$$\Gamma(2) = 1$$

If $\alpha = 1$ then it becomes exponential density

If $\lambda = 1$ then it is a one parameter gamma density

$$\begin{aligned}
 \int_{-\infty}^{\infty} \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t} dt &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} t^{\alpha-1} e^{-\lambda t} dt \\
 &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \int_0^{\infty} t^{\alpha-1} e^{-t/\frac{1}{\lambda}} dt \\
 &= \frac{\lambda^\alpha}{\Gamma(\alpha)} \Gamma(\alpha) \left(\frac{1}{\lambda}\right)^\alpha \\
 &= 1.
 \end{aligned}$$

Normal (Gaussian) distribution

→ Standard normal CDF = $\Phi(X)$

→ Standard normal density = $\phi(X)$

Beta Density

$$\text{Beta function: } B(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx \quad \text{or } B(m, n) = \int_0^{\infty} \frac{x^{n-1}}{(1+x)^{m+n}} dx$$

To prove density function:

For $0 < x < 1$ and $m > 0, n > 0$

$$\begin{aligned}
 \int_0^1 \frac{1}{B(m, n)} x^{m-1} (1-x)^{n-1} dx &= \frac{1}{B(m, n)} B(m, n) \\
 &= 1.
 \end{aligned}$$

$$\begin{aligned}
 B(m+1; n+1) &= \frac{m \Gamma(m) n \Gamma(n)}{(m+n+1) \Gamma(m+n+1)} \\
 &= \frac{mn}{(m+n+1)(m+n)} \times \frac{\Gamma(m) \Gamma(n)}{\Gamma(m+n)} \\
 &= \frac{mn}{(m+n+1)(m+n)} B(m; n).
 \end{aligned}$$

Different characteristics of the beta function			
Characteristic	Comment	Proof / Examples	Ref no
Type 1 beta function $\int_0^1 x^{m-1} (1-x)^{n-1} dx$ $= B(m; n)$	Recognize and use beta integral in integration	$\int_0^1 x^5 (1-x)^6$ $= \int_0^1 x^{6-1} (1-x)^{7-1}$ $= B(6; 7)$	B1
Symmetry with respect to the parameters $m > 0; n > 0$	$B(m; n) = B(n; m)$		B2
Relationship between beta and gamma functions	$m > 0; n > 0$ Numerical value for any beta function	$B(m; n) = \frac{\Gamma(m) \Gamma(n)}{\Gamma(m+n)}$	B4
Changing the limits of the integral from $\int_0^1 \dots$ to $\int_0^\infty \dots$	$\int_0^\infty \frac{x^{n-1}}{(1+x)^{m+n}} dx$ $= B(m; n)$ which is a type 2 beta function	$\int_0^\infty \frac{x^3}{(1+x)^5} dx$ $= \int_0^\infty \frac{x^{4-1}}{(1+x)^{4+1}} dx$ $= B(4; 1)$	

→ If $m = n = 1$ then becomes standard uniform distribution.

Cauchy Density

$$f(x) = \frac{1}{\pi} \left(\frac{1}{1+x^2} \right) \quad -\infty < x < \infty$$

Functions of random Variables

Proposition A:

if $X \sim N(\mu, \sigma^2)$ and $Y = aX + b$, then $Y \sim N(a\mu + b, a^2\sigma^2)$

$$\begin{aligned}
 P(x_0 < X < x_1) &= F_X(x_1) - F_X(x_0) \\
 &= \Phi\left(\frac{x_1 - \mu}{\sigma}\right) - \Phi\left(\frac{x_0 - \mu}{\sigma}\right)
 \end{aligned}$$

→ 2 Methods

- 1) Cumulative distribution function $F(x)$ method
 - find the **CDF** of the transformed variable
 - then differentiate to find the density function
 - and specify the region
- 2) Transformation method (Proposition B)

PROPOSITION B

Let X be a continuous random variable with density $f(x)$ and let $Y = g(X)$ where g is a differentiable, strictly monotonic function on some interval I . Suppose that $f(x) = 0$ if x is not in I . Then Y has the density function

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

for y such that $y = g(x)$ for some x , and $f_Y(y) = 0$ if $y \neq g(x)$ for any x in I . Here g^{-1} is the inverse function of g ; that is, $g^{-1}(y) = x$ if $y = g(x)$. ■

PROPOSITION C

Let $Z = F(X)$; then Z has a uniform distribution on $[0, 1]$.

Proof

$$P(Z \leq z) = P(F(X) \leq z) = P(X \leq F^{-1}(z)) = F(F^{-1}(z)) = z$$

This is the uniform cdf. ■

PROPOSITION D

Let U be uniform on $[0, 1]$, and let $X = F^{-1}(U)$. Then the cdf of X is F .

Proof

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x) \quad \blacksquare$$

Use D for generating pseudorandom numbers

$$[N(0, 1)]^2 \sim \chi_1^2$$

$$\text{Weibull density: } \frac{\beta}{\alpha^\beta} x^{\beta-1} e^{-\left(\frac{x}{\alpha}\right)^\beta}$$

Study Unit 3:: Joint Distributions**Discrete Random Variables**

$$p_{XY}(x, y) = P(X = x, Y = y)$$

$$1. \quad p_{X,Y}(x, y) \geq 0 \quad \text{for all } (x, y) \in R^2$$

$$2. \quad \sum_{x=0}^{n_1} \sum_{y=0}^{n_2} p_{X,Y}(x, y) = 1$$

Continuous Random Variables

$$1. \quad f_{X,Y}(x, y) \geq 0 \quad \text{for all } (x, y) \in R^2$$

$$2. \quad \int \int_S f_{X,Y}(x, y) dy dx = 1$$

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

Marginal density

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x;v) dv \quad \text{for all } x \in R$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(u;y) du \quad \text{for all } y \in R$$

Marginals sum to 1

Joint cumulative distribution.

$$\begin{aligned} F_{X,Y}(x;y) &= P(X < x; Y < y) \\ &= \int_{-\infty}^y \int_{-\infty}^x f_{X,Y}(u;v) dudv \end{aligned}$$

Independent Random variables

Independent if $F(x,y) = F_X(x)F_Y(y)$

Two discrete random variables will be independent if their joint mass function factors.

Conditional Distributions

Probabilities

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{if } P(B) \neq 0.$$

→ with independence

$$P(A|B) = \frac{P(A)P(B)}{P(B)} = P(A)$$

In the same way :

$$P(B|A) = \frac{P(A)P(B)}{P(A)} = P(B)$$

Discrete:

$$\begin{aligned} P(X = x_i | Y = y_j) &= \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} \\ &= \frac{p_{XY}(x_i, y_j)}{p_Y(y_j)} \end{aligned}$$

The law of total probability. $p_X(x) = \sum_y p_{X|Y}(x|y)p_Y(y)$

Continuous:

$$f_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

The law of total probability. $f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x)f_X(x) dx$

Functions of Jointly Distributed Random Variables

Discrete Convolution

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p_X(x)p_Y(z-x)$$

Continuous Convolution

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

Method of cumulative distribution functions

Composite variables	(x, y) coordinate specification	Graph to use	Distribution function of composite function U
$U = X + Y$	$x + y \leq u$	$x + y = u$ line $y = -x + u$	negative gradient y intercept u x intercept u
$U = X - Y$	$x - y \leq u$	$x - y = u$ line $y = x - u$	positive gradient y intercept $-u$ x intercept u
$U = \frac{X}{Y}$	$\frac{x}{y} \leq u$	$\frac{x}{y} = u$ line $y = \frac{1}{u}x$	gradient $= \frac{1}{u}$ Line passes through origin
$U = X \cdot Y$	$xy \leq u$	$xy = u$ hyperbola $y = \frac{u}{x}$	not defined in $y = 0$ and $x = 0$

Number of example	Composite variables	Given joint density with x and y regions	Distribution function of U
3.9	$U = \frac{X+Y}{2}$	$f(x, y) = e^{-x}e^{-y}$ $x > 0, y > 0$	$P\left(\frac{X+Y}{2} \leq u\right)$
3.10	$U = Y - X$	$f(x, y) = u$ $0 < x < y < \infty$	$P(Y - X \leq u)$
3.11	$U = \frac{X}{X+Y}$	$f(x, y) = xe^{-x}e^{-y}$ $x > 0, y > 0$	$P\left(\frac{X}{X+Y} \leq u\right)$
3.12	$U = X \cdot Y$	$f(x, y) = x + y$ $0 < x, y < 1$	$1 - P(XY \geq u)$
3.13	$U = X - Y$	$f(x, y) = 3x$ $0 < y < x < 1$	$1 - P(X - Y \geq u)$
3.14	$U = \frac{Y}{X}$	$f(x, y) = \frac{1}{8}xe^{-\frac{(x+y)}{2}}$ $x > 0, y > 0$	$P\left(\frac{Y}{X} \leq u\right)$

Composite variables	Change of variable y to u	Formula method Resulting formula
$U = X + Y$	$P(X + Y \leq u)$ $y = u - x$ $\frac{dy}{du} = 1$	$f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, u - x) dx$
$U = X - Y$	$P(X - Y \leq u)$ $y = x - u$ $\frac{dy}{du} = -1$	$f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}(x, x - u) dx$
$U = \frac{Y}{X}$	$P\left(\frac{Y}{X} \leq u\right)$ $y = ux$ $\frac{dy}{du} = x$	$f_U(u) = \int_{-\infty}^{\infty} x f_{X,Y}(x, ux) dx$
$U = X \cdot Y$	$P(XY \leq u)$ $y = \frac{u}{x}$ $\frac{dy}{du} = \frac{1}{x}$	$f_U(u) = \int_{-\infty}^{\infty} \frac{1}{ x } f_{X,Y}\left(x, \frac{u}{x}\right) dx$

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

Study Unit 4

Expected Value of Random Variables

Discrete $E(X) = \sum x p_X(x)$ Provided $\sum |x| p_X(x) < \infty$

Continuous $E(X) = \int_{-\infty}^{\infty} x f_X(x) dx$ Provided $\int_{-\infty}^{\infty} |x| f_X(x) dx < \infty$

Markov's Inequality

$$P(X \geq t) \leq \frac{E(X)}{t}$$

Expectations of functions of random variables

→ Corollary A: $E(XY) = E(X)E(Y)$

Expectations of linear combinations of random variables

THEOREM A

If X_1, \dots, X_n are jointly distributed random variables with expectations $E(X_i)$ and Y is a linear function of the X_i , $Y = a + \sum_{i=1}^n b_i X_i$, then

$$E(Y) = a + \sum_{i=1}^n b_i E(X_i)$$

→ Proof for $n = 2$:

- 1) Write out full expectation
- 2) Multiply out and separate integrals
- 3) First sums to 1
- 4) Second/third sum to expected values
- 5) State integral is convergent

Variance and standard deviation

$$\text{Var}(X) = E([X - E(X)]^2)$$

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

Discrete	$\text{Var}(X) = \sum_{x \in \mathcal{X}} (x - \mu)^2 p_X(x)$
Continuous	$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$
Standard deviation	$SD = \sqrt{\text{Var}(X)}$

Chebyshev's Inequality: There is a high probability that X will deviate little from its mean if the variance is small.

Let X be a random variable with mean μ and variance σ^2 . Then, for any $t > 0$,

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

Setting $t = k\sigma$ then $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$ or what ever is asked.

Covariance

$$\begin{aligned} \text{Cov}(X, Y) &= E(X - \mu_X)(Y - \mu_Y) \\ &= E(X, Y) - E(X)E(Y) \end{aligned}$$

$$\text{Cov}(X, Y) = \int \int (x - \mu_X)(y - \mu_Y) f_{X,Y}(x, y) dx dy$$

If independent then: $E(XY) = E(X)E(Y)$ and covariance is 0

THEOREM A

Suppose that $U = a + \sum_{i=1}^n b_i X_i$ and $V = c + \sum_{j=1}^m d_j Y_j$. Then

$$\text{Cov}(U, V) = \sum_{i=1}^n \sum_{j=1}^m b_i d_j \text{Cov}(X_i, Y_j)$$

COROLLARY A

$$\text{Var}(a + \sum_{i=1}^n b_i X_i) = \sum_{i=1}^n \sum_{j=1}^n b_i b_j \text{Cov}(X_i, X_j).$$

COROLLARY B

$$\text{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \text{Var}(X_i), \text{ if the } X_i \text{ are independent.}$$

$$\begin{aligned} \text{Var}(X + Y) &= \text{Cov}(X + Y, X + Y) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \end{aligned}$$

$$E\left(\sum X_i\right) = \sum E(X_i)$$

$$\text{Var}\left(\sum X_i\right) = \sum \text{Var}(X_i) \text{ if } X_i \text{ are independent.}$$

Correlation coefficient

Revise Properly!!!!!!

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

Conditional Expectation and Prediction

$$E(Y|X = x) = \sum_y y p_{Y|X}(y|x)$$

$$E(Y|X = x) = \int y f_{Y|X}(y|x) dy$$

$$E[h(Y)|X = x] = \int h(y) f_{Y|X}(y|x) dy$$

Law of total expectation

THEOREM A

$$E(Y) = E[E(Y|X)].$$

Proof

We will prove this for the discrete case. The continuous case is proved similarly. Using Theorem 4.1.1A we need to show that

$$E(Y) = \sum_x E(Y|X = x) p_X(x)$$

where

$$E(Y|X = x) = \sum_y y p_{Y|X}(y|x)$$

Interchanging the order of summation gives us

$$\sum_x E(Y|X = x) p_X(x) = \sum_y y \sum_x p_{Y|X}(y|x) p_X(x)$$

(It can be shown that this interchange can be made.) From the law of total probability, we have

$$p_Y(y) = \sum_x p_{Y|X}(y|x) p_X(x)$$

Therefore,

$$\sum_y y \sum_x p_{Y|X}(y|x) p_X(x) = \sum_y y p_Y(y) = E(Y) \quad \blacksquare$$

Moment-generating functions

$$M_X(t) = E(e^{tX})$$

discrete : $M_X(t) = \sum_x e^{tx} p_X(x)$

continuous : $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x) dx$

→ Same mgf then same distribution.

→ If the mgf can be determined it can be uniquely determines then probability distribution.

→ The mgf provides an elegant way to compute the moments of a distribution.

Calculating moments

→ First principles

→ Discrete: $E(X^r) = \sum x^r p_X(x)$

→ Continuous: $E(X^r) = \int x^r f_X(x) dx$

→ Using moment-generating function

$$\rightarrow M^{(r)}(0) = E(X^r)$$

→ Using Taylor series expansion

$$\begin{aligned} e^{tx} &= 1 + (tx) + \frac{(tx)^2}{2!} + \frac{(tx)^3}{3!} + \dots \\ &= 1 + (x) \cdot \frac{t}{1!} + (x)^2 \cdot \frac{t^2}{2!} + (x)^3 \cdot \frac{t^3}{3!} + \dots \end{aligned}$$

$$E(e^{tX}) = M_X(t) = 1 + E(X) \frac{t}{1!} + E(X^2) \frac{t^2}{2!} + E(X^3) \frac{t^3}{3!} + \dots$$

Properties of moment-generating functions

If X has the mgf $M_X(t)$ and $Y = a + bX$, then Y has the mgf $M_Y(t) = e^{at} M_X(bt)$.

If X and Y are independent random variables with mgf's M_X and M_Y and $Z = X + Y$, then $M_Z(t) = M_X(t)M_Y(t)$ on the common interval where both mgf's exist.

→ The joint moment-generating function of two random variables X and Y

$$M_{XY}(s, t) = E(e^{tX+sY}) = \sum_x \sum_y e^{tx+sy} p_{X,Y}(x, y) \text{ if } (X, Y) \text{ is discrete}$$

$$M_{XY}(s, t) = E(e^{tX+sY}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{tx+sy} f_{X,Y}(x, y) dx dy \text{ if } (X, Y) \text{ is continuous}$$

→ If X and Y are independent random variables then

$$M_{XY}(s, t) = M_X(s) M_Y(t)$$

Approximate problems

Go over examples in rice again

→ Propagation of error or δ method

→ The first order approximation of μ_X is

$$Y = g(X) \approx g(\mu_X) + (X - \mu_X) g'(\mu_X)$$

→ and the mean and the variance of Y are

$$\mu_Y \approx g(\mu_X) \quad \text{and} \quad \sigma_Y^2 \approx \sigma_X^2 [g'(\mu_X)]^2.$$

→ The second order approximation of μ_X is

$$Y = g(X) \approx g(\mu_X) + (X - \mu_X) g'(\mu_X) + \frac{1}{2} (X - \mu_X)^2 g''(\mu_X)$$

↑ Expected Value = 0

→ and the new improved mean of Y is

$$E(Y) \approx g(\mu_X) + \frac{1}{2} \sigma_X^2 g''(\mu_X).$$

Study Unit 5

Law of large numbers

THEOREM A *Law of Large Numbers*

Let $X_1, X_2, \dots, X_i \dots$ be a sequence of independent random variables with $E(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Let $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$. Then, for any $\varepsilon > 0$,

$$P(|\bar{X}_n - \mu| > \varepsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

Proof

We first find $E(\bar{X}_n)$ and $\text{Var}(\bar{X}_n)$:

$$E(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \mu$$

Since the X_i are independent,

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}$$

The desired result now follows immediately from Chebyshev's inequality, which states that

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0, \quad \text{as } n \rightarrow \infty \quad \blacksquare$$

Central limit Theorem

→ Remember approximation \approx and not $=$ when doing calculations because only equal in the limit.

Let X_1, X_2, \dots be a sequence of independent random variables having mean 0 and variance σ^2 and the common distribution function F and moment-generating function M defined in a neighborhood of zero. Let

$$S_n = \sum_{i=1}^n X_i \quad \text{no need to prove}$$

Then

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), \quad -\infty < x < \infty$$

Study Unit 6

χ^2 , t and F distributions

→