

STA2603 Distribution Theory II

Always specify the domain of the function!

Always specify the parameters when identifying the distribution!

Study Unit 2

Frequency Function (discrete) = density function (continuous).

Distribution functions is CDF.

Discrete Random Variables

Distribution	Mass function $p_X(k)$	$M_X(t)$ mgf	Mean & Variance
Bernoulli $ber(p)$ one success	$\begin{cases} p^k (1-p)^{1-k} & k = 0; 1 \\ 0 & \text{elsewhere} \end{cases}$	$(1-p + pe^t)$	$\begin{aligned} \mu_X &= p \\ \sigma_X^2 &= p(1-p) \end{aligned}$
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	$\begin{cases} (1-p)^{x-1} p, & x = 1; 2; 3; \dots \\ 0 & \text{elsewhere} \end{cases}$	$\frac{pe^t}{1-(1-p)e^t}$	$\begin{aligned} \mu_X &= \frac{1}{p} \\ \sigma_X^2 &= \frac{1-p}{p^2} \end{aligned}$
Negative Binomial $nb(r; p)$ r'th success in k trials	$\begin{cases} \binom{k-1}{r-1} p^r (1-p)^{k-r}, & k = 1; 2; \dots \\ 0 & \text{elsewhere} \end{cases}$	$\left[\frac{pe^t}{1-(1-p)e^t} \right]^r$	$\begin{aligned} \mu_X &= \frac{r}{p} \\ \sigma_X^2 &= \frac{r(1-p)}{p^2} \end{aligned}$
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. $\sum_{x=0}^{\infty} p_X(x) = 1$

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

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:

$$\begin{aligned} (a+b)^n &= a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \dots + \binom{n}{n-1} a^{n-(n-1)}b^{n-1} + \binom{n}{n} a^{n-n}b^n \\ &= (a+b)^n = a^n + \binom{n}{1} a^{n-1}b + \binom{n}{2} a^{n-2}b^2 + \dots + b^n \\ &= \sum_{x=0}^n \binom{n}{x} a^{n-x}b^x \end{aligned}$$

and therefore it is a series expansion of $[(1-p) + p]^n$

Geometric Distribution

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

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

Negative Binomial Distribution

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

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:

$$e^{\lambda} = 1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \frac{\lambda^4}{4!} + \dots$$

$$= \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}$$

$$\sum_{x=0}^{\infty} \frac{\lambda^k}{x!} e^{-\lambda} = e^{\lambda} e^{-\lambda} = 1.$$

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

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

1. $f_X(x) \geq 0$ and

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

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

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

Exponential Density

Gamma Density

Gamma function $\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx$ Properties $\Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$ $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

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

$$\Gamma(2) = 1$$

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

Beta Density

Beta function: $B(m, n) = \int_0^1 x^{m-1} (1-x)^{n-1} dx$ 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	$\int_0^\infty \frac{x^3}{(1+x)^5} dx$ $= \int_0^\infty \frac{x^{4-1}}{(1+x)^{4+1}} dx$ $= B(4; 1)$	

Cauchy Density

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

Functions of random Variables

Standard normal CDF = $\Phi(X)$ Standard normal density = $\phi(X)$

Proposition A:

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

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

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

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

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

Study Unit 3**Discrete Random Variables**

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

1. $p_{X,Y}(x, y) \geq 0$ for all $(x, y) \in R^2$
2. $\sum_{x=0}^{n_1} \sum_{y=0}^{n_2} p_{X,Y}(x, y) = 1$

Continuous Random Variables

Marginal density.

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

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

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

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

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

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

Discrete:

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

Continuous:

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

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

$$Var(X) = E([X - E(X)]^2)$$

$$Var(X) = E(X^2) - [E(X)]^2$$

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

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

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

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

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

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

COROLLARY A

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

COROLLARY B

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

$$\begin{aligned}\text{Var}(X + Y) &= \text{Cov}(X + Y, X + Y) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X, Y) \\ E\left(\sum X_i\right) &= \sum E(X_i) \\ \text{Var}\left(\sum X_i\right) &= \sum \text{Var}(X_i) \text{ if } X_i \text{ are independent.}\end{aligned}$$

Correlation coefficient

Revise Properly!!!!!!

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

Conditional Expectation and Prediction

$$\begin{aligned}E(Y|X = x) &= \sum_y y p_{Y|X}(y|x) \\ E(Y|X = x) &= \int y f_{Y|X}(y|x) dy \\ E[h(Y)|X = x] &= \int h(y) f_{Y|X}(y|x) dy\end{aligned}$$

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

- Using moment-generating function

→ $M^{(r)}(0) = E(X^r)$

- Using Taylor series expansion

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

Study Unit 5

Study Unit 6