

Notes of Probability, MATH 2005

Box, ZHANG Huakang

January 5, 2021

Contents

1	Probability Measure	2
2	Conditional Probability	2
3	Discrete Uniform Distributions	3
4	Bernoulli Distributions	3
5	Binomial Distributions	4
6	Negative Binomial Distributions	5
7	Geometric Distributions	6
8	Hyper-geometric Distributions	6
9	Possion Distributions	7
10	Multivariate Distributions	7
10.1	Polynoms	7
10.2	Multivariate Hyper-geometric Distributions	8
11	Uniform Densities	8
12	The Gamma Function	9
13	The Beta Function	9
14	Gamma Distributions	9
15	Beta Distributions	10
16	Exponential Distributions	11
17	Normal Distributions	11
17.1	The Standard Normal Distributions	11
17.2	Normal Distributions	11
17.3	The Normal Approximation	12
18	Bivariate Normal Distributions	12
19	The Distribution Function Technique	13
20	The Transformation Technique	13

21 Moment Generating Functions	13
21.1 Some Moment-generating functions	13
21.1.1 Binomial random variable	13
21.1.2 Poisson random variable	14
21.1.3 Gamma random variable	14
21.1.4 Exponential random variable	14
21.1.5 Chi-square random variable	14
21.1.6 Normal random variable	14
21.2 Properties of Moment Generating Functions	14
22 Moment Generating Function Technique	15
22.1 Poisson Distribution	15
22.2 Normal Distribution	15
22.3 Exponential Distribution	15

1 Probability Measure

A probability measure must satisfy the following three postulates:

1. $\mathbb{P}(S) = 1$
2. For each event A , the probability of A is a nonnegative real number, i.e., $\mathbb{P}(A) \geq 0$
3. if $\{A_n\}$ is an infy sequence of events if F such that, for any $i \neq j$, $A_i \cap A_j = \emptyset$, then

$$\mathbb{P}(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbb{P}(A_n)$$

2 Conditional Probability

Let (S, F, \mathbb{P}) be a probability space, and let A and B are two random events in the sample space S with $\mathbb{P}(B) \neq 0$. Then the Conditional probability of A given B is defined by

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

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

Let n random events

3 Discrete Uniform Distributions

Definition

A discrete random variable X is said to have a *discrete uniform distribution*, and it is called a *discrete uniform variable*, if it can take on k different values: x_1, x_2, \dots, x_k , and its probability distribution $f(x)$ is given by

$$f(x_i) = \frac{1}{k}$$

where $i = 1, 2, \dots, k$.

Mean and Variance

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=1}^k x_i f(x_i) \\ &= \frac{1}{k} \sum_{i=1}^k x_i \\ \mathbb{E}[X^2] &= \sum_{i=1}^k x_i^2 f(x_i) \\ &= \frac{1}{k} \sum_{i=1}^k x_i^2 \\ \text{var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \frac{1}{k} \sum_{i=1}^k x_i^2 - \left(\frac{1}{k} \sum_{i=1}^k x_i\right)^2\end{aligned}$$

4 Bernoulli Distributions

Definition

Support that an experiment has two possible outcomes: success and failure, and their probability are respectively, θ and $1 - \theta$. Then, this experiment is called a *Bernoulli Distributions*. Let X be the number of successes of a Bernoulli experiment, i.e. $X = 1$ or $X = 0$. Then, X is called a random variable having the Bernoulli probability distribution, which is given by

$$f(x; \theta) = \theta^x (1 - \theta)^{1-x}$$

where $x = 0, 1$ and $0 < \theta < 1$ is a parameter.

Mean and Variance

$$\begin{aligned}\mathbb{E}[X] &= \theta \\ \mathbb{E}[X^2] &= \theta \\ \text{var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \theta - \theta^2 \\ &= \theta(1 - \theta)\end{aligned}$$

5 Binomial Distributions

Definition

Let n be a natural number, and let $0 < \theta < 1$. Then, a discrete random variable X is said to have a *binomial distribution*, and X is called a binomial random variable, if its probability distribution $b(x; n, \theta)$ is given by

$$b(x; n, \theta) = C_n^x \theta^x (1 - \theta)^{n-x}$$

where $x = 1, 2, \dots, n$ and θ are two parameters, and

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

is the total number of combinations of n distinct numbers taken x numbers at a time.

Remark

We consider n independent Bernoulli experiments, in which the parameter θ (the probability of a success) is the same for each experiment. Let X be the total number of successes in this sequence of n independent Bernoulli experiments. Then, we can see that X is a random variable having a binomial distribution with parameters n and θ , i.e., we have the following result.

Let X_1, X_2, \dots, X_n be n independent Bernoulli random variables with the same parameter θ . Then, the random variable $X = X_1 + X_2 + \dots + X_n$ has a binomial distribution with parameters n and θ .

Mean and Variance

$$\begin{aligned}\mathbb{E}[X] &= n\theta \\ \text{var}(X) &= \text{var}(X_1 + X_2 + \dots + X_n) \\ &= \text{var}(X_1 + X_2 + \dots + X_{n-1}) + \text{var}(X_n) - 2\text{cov}(X_1 + X_2 + \dots + X_{n-1}, X_n) \\ &\quad \dots \\ &= \text{var}(X_1) + \text{var}(X_2) + \dots + \text{var}(X_n) \\ &= n\theta(1 - \theta)\end{aligned}$$

Theorem

$$\begin{aligned} b(x; n, \theta) &= C_n^x \theta^x (1 - \theta)^{n-x} \\ &= C_n^{n-x} (1 - \theta)^{1-\theta} \theta^x \\ &= b(n - x; n, 1 - \theta) \end{aligned}$$

Since a binomial random variable X with parameters n and θ is the total number of successes in n independent Bernoulli experiments. $Y = \frac{X}{n}$ is the proportion of successes in n independent Bernoulli experiments.

$$\mathbb{E}[Y] = \theta$$

$$\text{var}(Y) = \frac{\theta(1 - \theta)}{n}$$

6 Negative Binomial Distributions

Definition

Let k be a natural number and let $0 < \theta < 1$. Then, a discrete random variable Y is said to have a (Pascal) negative binomial distribution, and it is called a (Pascal) negative binomial random variable, if its probability distribution $b^*(y; k, \theta)$ is given by

$$b^*(y; k, \theta) = C_{y-1}^{k-1} \theta^k (1 - \theta)^{y-k}$$

where k and θ are two parameters.

Mean and Variance

$$\begin{aligned} \mathbb{E}[Y] &= \sum_{i=k}^{\infty} i b^*(i; k, \theta) \\ &= \sum_{i=k}^{\infty} i C_{i-1}^{k-1} \theta^k (1 - \theta)^{i-k} \\ &\quad \dots \\ &= \frac{k}{\theta} \\ \text{var}(Y) &= \frac{k}{\theta} \left(\frac{1}{\theta} - 1 \right) \end{aligned}$$

Theorem

Let Y be a negative binomial random variable with parameters k and θ . Then for each $y = k, k + 1, \dots$,

$$b^*(y; k, \theta) = \frac{k}{y} b(k; y, \theta)$$

Proof. By the definition, we have

$$\begin{aligned}
b^*(y; k, \theta) &= C_{y-1}^{k-1} \theta^k (1 - \theta)^{y-k} \\
&= \frac{(y-1)!}{(k-1)!(y-k)!} \theta^k (1 - \theta)^{y-k} \\
&= \frac{k}{y} \frac{y!}{k!(y-k)!} \theta^k (1 - \theta)^{y-k} \\
&= \frac{k}{y} b(k; y, \theta)
\end{aligned}$$

□

7 Geometric Distributions

Definition

If X is a (Pascal) negative binomial random variable with parameters $k = 1$ and θ , we say that this random variable X has a geometric distribution, and we also call this random variable as a geometric random variable. By the definition of negative binomial distribution, we see that the probability distribution $g(x; \theta) = b^*(x; 1, \theta)$ of geometric distribution is given by

$$g(x; \theta) = \theta(1 - \theta)^{x-1}$$

where θ is a parameter.

Mean and Variance

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{\theta} \\
\text{var}(X) &= \frac{1}{\theta} \left(\frac{1}{\theta} - 1 \right)
\end{aligned}$$

Theorem

The geometric distribution has the memoryless property, i.e. if X is a geometric random variable, then, for any nature n ,

$$\mathbb{P}(X = x + n | X > n) = \mathbb{P}(X = x)$$

8 Hyper-geometric Distributions

Definition

A random variable X is said to have a hyper-geometric distribution, and it is referred to as a hyper-geometric random variable, if its

probability distribution is given by

$$h(x; n, N, k) = \frac{C_k^x C_{N-k}^{n-x}}{C_N^n}$$

for $x = 0, 1, \dots, n$ with $x \leq k$ and $n - x \leq N - k$, where n, N, k are parameters.

Mean and Variance

$$\begin{aligned}\mathbb{E}[X] &= \frac{nk}{N} \\ \text{var}(X) &= \frac{nk(N-k)(N-n)}{N^2(N-1)}\end{aligned}$$

9 Poisson Distributions

Definition

A discrete random variable X is said to have a Poisson distribution, and it is referred to as a Poisson random variable, if its probability distribution is given by

$$p(x; \lambda) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Mean and Variance

$$\begin{aligned}\mathbb{E}[X] &= \lambda \\ \text{var}(X) &= \lambda\end{aligned}$$

10 Multivariate Distributions

10.1 Polynomial Distributions

Definition

The k random variable X_1, X_2, \dots, X_k are said to have a polynomial distribution, and they are referred to as polynomial random variable, if their joint probability distribution is given by

$$f(x_1, \dots, x_k; \theta_1, \dots, \theta_k) = \frac{n!}{x_1! \dots x_k!} \theta_1^{x_1} \dots \theta_k^{x_k}$$

for $x_i = 0, 1, \dots, n$ and $0 < \theta_i < 1$ for each $i = 1, \dots, k$, where

$$n = \sum_{i=1}^k x_i$$

$$\sum_{i=1}^k \theta_i = 1$$

10.2 Multivariate Hyper-geometric Distributions

Definition

The k random variable X_1, \dots, X_k are said to have a multivariate hyper-geometric distribution, and they are referred to as multivariate hyper-geometric random variable, if their joint probability distribution is given by

$$f(x_1, \dots, x_k; n, j_1, \dots, j_k) = \frac{C_{j_1}^{x_1} \dots C_{j_k}^{x_k}}{C_N^n}$$

for $x_i = 0, 1, \dots, n$ with $x_i \leq j_i$ for each $i = 1, \dots, k$, where

$$n = \sum_{i=1}^k x_i$$

$$N = \sum_{i=1}^k j_i$$

11 Uniform Densities

Definition

A continuous random variable X is said to have a uniform density with parameters α and β , and it is referred to as a continuous uniform random variable, if its probability density is given by

$$f(x) = \frac{1}{\beta - \alpha}$$

for $\alpha < x < \beta$, and $f(x) = 0$ elsewhere, where α and β are two parameters with $\alpha < \beta$.

Mean and Variance

$$\mathbb{E}[X] = \frac{\alpha + \beta}{2}$$

$$\text{var}(X) = \frac{(\beta - \alpha)^2}{12}$$

12 The Gamma Function

Definition

The gamma function $\Gamma(x)$ is a real function which is defined on $(0, \infty)$ and is given by

$$\Gamma(x) = \int_0^\infty y^{x-1} e^{-y} dy$$

for each $x \in (0, \infty)$

Theorem

$$\Gamma(n) = (n-1)!$$

for each natural number n .

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

13 The Beta Function

Definition

The beta function $B(u, v)$ is a real bivariate function, which is defined for each $(u, v) \in (0, \infty) \times (0, \infty)$, and is given by

$$B(u, v) = \int_0^1 t^{u-1} (1-t)^{v-1} dt$$

Theorem

The gamma function and the beta function satisfy the following equation

$$B(u, v) = \frac{\Gamma(u)\Gamma(v)}{\Gamma(u+v)}$$

for any $(u, v) \in (0, \infty) \times (0, \infty)$

14 Gamma Distributions

Definition

A random variable X is said to have a gamma distribution, and it is referred to as a gamma random variable, if its probability density is

given by

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}$$

for $x > 0$, and $f(x) = 0$ for $x \leq 0$, where $\alpha > 0$ and $\beta > 0$ are two parameters.

Theorem

The r -th moment of a gamma distribution with parameters α and β is given by

$$\mu'_r = \frac{\beta^r \Gamma(\alpha + r)}{\Gamma(\alpha)}$$

Mean and Variance

$$\begin{aligned}\mathbb{E}[X] &= \alpha\beta \\ \text{var}(X) &= \alpha\beta^2\end{aligned}$$

15 Beta Distributions

Definition

A random variable X is said to have a beta distribution, and it is referred to as a beta random variable, if its probability density is given by

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}$$

Theorem

The r -th moment of a beta distribution with parameters α and β is given by

$$\mu'_r = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)} \frac{\Gamma(\alpha + r)}{\Gamma(\alpha + \beta + r)}$$

Mean and Variance

$$\begin{aligned}\mathbb{E}[X] &= \frac{\alpha}{\alpha + \beta} \\ \text{var}(X) &= \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}\end{aligned}$$

16 Exponential Distributions

Definition

Let X be a gamma random variable with parameters α and β . If $\alpha = 1$ and $\beta = \theta$, this random variable X is said to have an exponential distribution, and it is referred to as an exponential random variable. The density of exponential distribution is given by

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$$

for $x > 0$, and $f(x) = 0$ for $x \leq 0$

Theorem

The exponential distribution satisfies the memoryless property, i.e., if X is an exponential random variable, then, for each $t > 0$,

$$\mathbb{P}(X \geq x + t | X \geq t) = \mathbb{P}(X \geq x)$$

Mean and Variance

$$\begin{aligned}\mathbb{E}[X] &= \theta \\ \text{var}(X) &= \theta^2\end{aligned}$$

17 Normal Distributions

17.1 The Standard Normal Distributions

Definition

A random variable Z is said to have the standard normal distribution, and so that it is called a standard normal distribution variable and it is denoted by $Z \sim N(0, 1)$, if its density is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

17.2 Normal Distributions

Definition

Let X be a normal random variable with parameters μ and $\sigma > 0$, and let $Z = \frac{X - \mu}{\sigma}$. Then, Z is a standard normal random variable, i.e., $Z \sim N(0, 1)$

Mean and Variance

$$\begin{aligned}\mathbb{E}[X] &= \mu \\ \text{var}(X) &= \sigma^2\end{aligned}$$

17.3 The Normal Approximation

Let X_n be a random variable having a binomial distribution with parameters n and θ , and let

$$Z_n = \frac{X_n - n\theta}{\sqrt{n\theta(1-\theta)}}$$

18 Bivariate Normal Distributions

Definition

A pair of random variables (X_1, X_2) is said to have a bivariate normal distribution, and X_1 and X_2 are referred to as jointly normal distributed random variables, if their joint density is given by

$$\Phi(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\psi(x_1, x_2)}$$

for $(x_1, x_2) \in \mathbb{R}^2$ where

$$\psi(x_1, x_2) = \frac{1}{2(1-\rho^2)} \left(\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right)$$

and $\mu_i, \sigma_i > 0$ with $i = 1, 2$, and $|\rho| < 1$ are all parameters.

Theorem

The marginal densities are respectively given by

$$\begin{aligned}\phi_1(x_1) &= \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x_1 - \mu_1)^2}{2\sigma_1^2}} \\ \phi_2(x_2) &= \frac{1}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x_2 - \mu_2)^2}{2\sigma_2^2}}\end{aligned}$$

The covariance of X_1 and X_2 is given by

$$\text{cov}(X_1, X_2) = \rho\sigma_1\sigma_2$$

X_1 and X_2 are independent, if and only if $\rho = 0$

19 The Distribution Function Technique

Let X be a random variable with the distribution function $F(x)$, and let $Y = u(X)$, where $u(x)$ is an increasing function such that its inverse function $x = u^{-1}(y)$ exists. Then the distribution function of Y is given by

$$G(y) = F(u^{-1}(y))$$

for all real numbers y in the range of Y .

20 The Transformation Technique

Let X be a random variable whose density function is $f(x)$, and let $y = u(x)$ is a differentiable function such that its inverse function $x = u^{-1}(y)$ exists. Then, the density function $g(y)$ of $Y = u(X)$ is given by

$$g(y) = f(u^{-1}(y))|u'(y)|$$

when $u'(y) \neq 0$ and $g(y) = 0$ elsewhere where $u(y) = u^{-1}(y)$.

21 Moment Generating Functions

Definition

Let X be a random variable. The moment-generating function of X is defined by

$$M_X(t) = \mathbb{E}[e^{tX}]$$

for each real number t in which the expectation exists, i.e., when X is continuous with density $f(x)$.

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

when X is discrete with probability distribution $f(x)$,

$$M_X(t) = \sum_x e^{tx} f(x)$$

21.1 Some Moment-generating functions

21.1.1 Binomial random variable

$$M_X(t) = (1 + \theta(e^t - 1))^n$$

21.1.2 Poisson random variable

$$M_X(t) = e^{\lambda(e^t - 1)}$$

21.1.3 Gamma random variable

$$M_X(t) = (1 - \beta t)^{-\alpha}$$

21.1.4 Exponential random variable

$$M_X(t) = \frac{1}{1 - \theta t}$$

21.1.5 Chi-square random variable

$$M_X(t) = (1 - 2t)^{-\frac{\nu}{2}}$$

21.1.6 Normal random variable

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

21.2 Properties of Moment Generating Functions

$$\frac{d^k M_X(t)}{dt^k} \Big|_{t=0} = \mu'_k = \mathbb{E}[X^k]$$

$$M_{aX+b}(t) = e^{tb} M_X(at)$$

Let X and Y be two independent random variables,

$$M_{X+Y}(t) = M_X(t)M_Y(t)$$

22 Moment Generating Function Technique

22.1 Poisson Distribution

Let X_1 and X_2 be two independent random variables, and let X_1 and X_2 have the Poisson distributions $p(x_1, \lambda_1)$ and $p(x_2, \lambda_2)$, respectively. Then, the random variable $Y = X_1 + X_2$ has a Poisson distribution

$$p(y, \lambda_1 + \lambda_2)$$

22.2 Normal Distribution

Let X_1 and X_2 be two independent random variables such that $X_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2$, respectively, and let a and b be two constants such that $a^2 + b^2 \neq 0$. then, the random variable $Y = aX_1 + bX_2$ has a normal distribution with mean $\mu = a\mu_1 + b\mu_2$ and variance $\sigma^2 = a^2\sigma_1^2 + b^2\sigma_2^2$.

22.3 Exponential Distribution

Let X_1 and X_2 be two independent random variables having an exponential distributions with the same parameter θ , respectively. Then the random variable $Y = X_1 + X_2$ has a gamma distribution with $\alpha = 2$ and $\beta = \theta$