AMS 572 Data Analysis I
Review of Probability

Silvia Sharna

Applied Math and Stats, Stony Brook University

Review of Probability

Discrete random variable

A random variable that can take at most a countable number of possible values is said to be discrete. The space of discrete random variable contains at most a countable number of points. If the space of a random variable contains an interval, the random variable is called continuous random variable.

For a discrete random variable X, we define the probability mass function (p.m.f. or prob distribution function, p.d.f.) $p(x)$ of X by

$$p(x) = P(X = x)$$

Cumulative distribution function

The cumulative distribution function (cdf) F of the random variable X is defined by $F(x) = P(X \leq x)$ for $-\infty < x < \infty$. For a discrete random variable X:

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} p(x_i)$$

It is a non decreasing step function. That is, if the possible values of X are $x_1, x_2, \ldots$ where $x_1 < x_2 < \ldots$, the F is constant in the intervals (x_{i-1}, x_i) and then takes a step (or jump) of size $p(x_i)$ at x_i.

Definition

If X is a discrete random variable having p.d.f. $p(x)$, the expectation or the expected value of X, $E(X)$ is defined by

$$E(X) = \sum_i x_i p(x_i)$$

Also known as mean μ : weighted average of the possible values of X.

Properties:

1. $E(c) = c$, c : constant.
2. $E(cU(X)] = cE(U(X))$.
3. $E[c_1U_1(X) + c_2U_2(X)] = c_1E(U_1(X)) + c_2E(U_2(X))$.

Definition

If X is a random variable with mean μ, then the variance of X, $\mathrm{Var}(X)$ is defined by

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \sum_i (x_i - \mu)^2 p(x_i)$$

which is the mean squared deviation with respect to μ.

Properties

1. $\mathrm{Var}(c) = 0$, if c is constant.
2. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$, a, b constant.
3. $\mathrm{Var}(X) = E[(X - \mu)^2] = E(X^2) - [E(X)]^2$.

Definition

The standard deviation of X is

$$\mathrm{sd}(X) = \sqrt{\mathrm{Var}(X)}$$

Binomial Distribution

Binomial experiment is one that possesses the following properties

1. A Bernoulli (success-failure) trial is performed n times.
2. The trials are independent
3. The probability of success on a single trial is equal to p and remains the same from trial to trial. The probability of failure is $(1 - p) = q$.
4. The random variable of interest is X : the number of successes observed during the n trials.

X is called the Binomial random variable. The p.d.f. of X is called Binomial distribution and denoted as

$$X \sim Bin(n, p)$$

The p.d.f. of X is

$$P(X = x) = p(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

for $x = 0, 1, \ldots, n$

Properties

1. $P(X = x) = p(x)$ is a p.d.f..
2. $E(X) = np$.
3. $\text{Var}(X) = np(1 - p)$.
4. When x goes from 0 to n, $p(x)$ first increases monotonically then decreases monotonically reaching its largest value when x is the largest integer less than or equal to $(n + 1)p$.

Example: There are three coins in a bag. When coin 1 is flipped, it lands on head with probability 0.3. When coin 2 is flipped, it lands on head with probability 0.8. When coin 3 is flipped, it lands on head with probability 0.6. One of these coins is randomly chosen and flipped 8 times.

(a) What is the probability that the coin lands on head exactly 3 out of the 8 flips?

(b) Given that the last 3 of these 8 flips lands on head, what is the conditional probability that exactly 6 out of the 8 flips lands on head?

Poisson Distribution

A random variable X taking on one of the values $0, 1, 2, \ldots$ is said to be a Poisson random variable with parameter λ if for some $\lambda > 0$, the p.d.f. of X is

$$P(X = x) = p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \ x = 0, 1, 2, \ldots$$

X is said to have a Poisson distribution with parameter λ and is denoted as $X \sim \text{Poisson}(\lambda)$ or $X \sim Po(\lambda)$ or $X \sim P(\lambda)$.

Properties of $X \sim P(\lambda)$.

1. $E(X) = \lambda$.
2. $\mathrm{Var}(X) = \lambda$.
   Note that $E(X) = \mathrm{Var}(X) = \lambda$ for a Poisson distribution.
3. If $X \sim P(\lambda)$, then $P(X = x)$ increases monotonically and then decreases monotonically as x increases, reaching maximum when x is the largest integer not exceeding λ.

When $n \geq 20$ and $p \leq 0.05$, the Poisson distribution usually gives a good approximation to the Binomial distribution.

Example: Suppose that on average 1 person in 1000 make a numerical error in preparing his/her income tax return. If 10000 people are selected at random and examined, find an approximation for the probability that 6, 7, or 8 of them contain an error.

Continuous Random Variable

Random variable whose set of possible values is an interval of union of intervals is said to be a continuous random variable (or a random variable of the continuous type).

The p.d.f. $f(x)$ is used to describe probability of events concerning the continuous random variable and satisfies the following conditions:

1. $f(x) \geq 0$.
2. $\int_R f(x)\,\mathrm{d}x = 1$, where R is the space of X.

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$$

3. The probability of event A is $\int_A f(x)\,\mathrm{d}x = P(X \in A)$

The c.d.f. of random variable X is given by

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

Expectation of X is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

Expectation of $U(X)$ is

$$\mu = E[U(X)] = \int_{-\infty}^{\infty} U(x) f(x)\,\mathrm{d}x$$

Moment generating function

Definition:

The moment generating function $M(t)$ of random variable of X is defined by

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

for $t \in \mathbb{R}$.

$M(t)$ gets its name because all the moments of X can be obtained by successfully differentially $M(t)$ and then evaluating the result at $t = 0$.

Moment generating function of X:

$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,\mathrm{d}x$$

1. $M'(0) = E(X)$
2. $M''(0) = E(X^2)$
3. $\mathrm{Var}(X) = M''(0) - [M'(0)]^2$

Normal Distribution

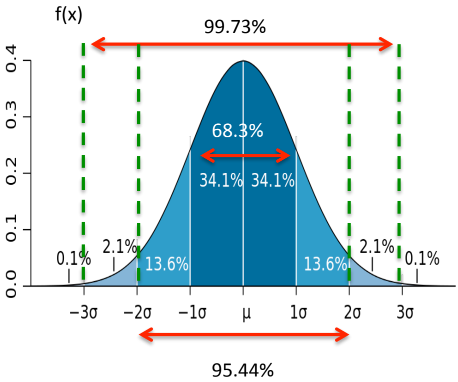
Normal distribution is the most important distribution in statistical applications because in practice, many measurements obey, at least approximately a normal distribution. The central limit theorem gives a theoretical base to this fact.

X is distributed as a normal random variable if its p.d.f. is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, -\infty < x < \infty$$

where σ^2 and μ are the parameters. Denote

$$X \sim N(\mu, \sigma^2)$$



$Z = (X - \mu)/\sigma \sim N(0, 1)$ Denote the c.d.f. of standard normal by $\Phi(x)$.

$$\Phi(x) = P(Z \le x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2}\, dt$$

1. Moment generating function.
   Consider $Z \sim N(0,1)$ first

$$\begin{aligned} M(t) &= E(e^{tZ}) \\ &= \int_{-\infty}^{\infty} e^{tz} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, \mathrm{d}z \end{aligned}$$

Distribution of a function of a random variable

Many problems require finding the distribution of some function of X, say $Y = g(X)$ from the distribution of X. Suppose X has density $f_X(x)$, where a subscript now is used to distinguish densities of different random variables. We want to find $f_Y(y)$.

Method I: c.d.f. Method (works for every transformation)
Example:
If X is a continuous random variable with p.d.f. f_X and $Y = X^2$, find $f_Y(y)$.

Example: Let $X \sim N(0,1)$. Find the p.d.f. of $Y = X^2$.

Method II: Jacobian Method

Let X be a random variable with density $f_X(x)$ on the range (a, b). Let $Y = g(X)$ where g is either strictly increasing of strictly decreasing on (a, b). Range of Y is an interval with endpoints $g(a)$ and $g(b)$. Then

$$f_Y(y) = f_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|$$

Example: Let $X \sim \text{Exp}(1)$, i.e., $f_X(x) = e^{-x}, x > 0$.
Find the density of $Y = e^{-X} = g(X)$.

Try Yourself

Read Chapters 5 and 6