STA2001 Probability and Statistics (I)

Lecture 10

Tianshi Chen

The Chinese University of Hong Kong, Shenzhen

Review

Definition

A RV X with $\overline{S}$ that is an interval or unions of intervals is said to be continuous RV. If there exists a function $f(x) : \overline{S} \to (0, \infty)$ such that

1. $f(x) > 0, \quad x \in \overline{S}$
2. $\int_{\overline{S}} f(x)dx = 1$
3. If $(a, b) \subseteq \overline{S}$

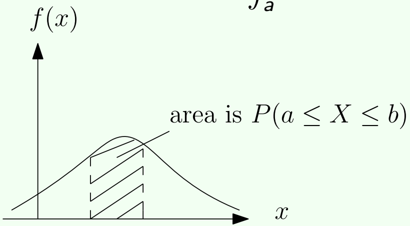
$$P(a \leq X \leq b) \triangleq \int_a^b f(x)dx$$

f is the so-called probability density function (pdf).

Review

Interpretation

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$



Review

Definition

cdf $F(x) : R \to [0,1]$

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

1. relation between the pdf and the cdf

$$f(x) = F'(x)$$

   for those values of x at which $F(x)$ is differentiable

2. relation between the probability function and the cdf

$$P(a \leq X \leq b) = F(b) - F(a)$$

Mathematical Expectation

Mathematical Expectation

Let X be a continuous RV with pdf $f(x) : \overline{S} \to (0, \infty)$. If $\int_{\overline{S}} g(x)f(x)dx$ exists, it is called the mathematical expectation for $g(X)$ and denoted by

$$E[g(X)] = \int_{\overline{S}} g(x)f(x)dx$$

If the range of X is extended from $\overline{S}$ to R with $f(x) = 0$ for $x \notin \overline{S}$, then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx$$

Expectation is a linear operator [Theorem 2.2-1, page 60].

$$E[c_1 g_1(X) + c_2 g_2(X)] = c_1 E[g_1(X)] + c_2 E[g_2(X)]$$

Special Mathematical Expectations

1. $[g(X) = X]$: Mean of X, $E[X] = \int_{\overline{S}} xf(x)dx$
2. $[g(X) = (X - E[X])^2]$: Variance of X,

$$Var[X] = E[(X - E[X])^2] = \int_{\overline{S}} (x - E[X])^2 f(x)dx$$

3. $[g(X) = X^r]$, Moments of X:

$$E[X^r] = \int_{\overline{S}} x^r f(x)dx$$

Special Mathematical Expectations

4. $[g(X) = e^{tX}]$: Moment generating function (mgf). If there exists $h > 0$, such that

$$M(t) = E[e^{tX}] = \int_S e^{tx} f(x)dx, \quad -h < t < h \text{ for some } h > 0$$

Mgf determines the distribution of X and all moments exist and are finite

$$M^{(r)}(0) = E[X^r]$$

which can be used to derive the mean and variance of a RV X

$$E[X] = M'(0), \quad Var[X] = M''(0) - (M'(0))^2$$

Example 3, page 98

Let X have the pdf

$$f(x) = \begin{cases} \frac{1}{100}, & 0 < x < 100 \\ 0, & \text{otherwise.} \end{cases} \Leftrightarrow X \sim U(0, 100)$$

Example 3, page 98

Let X have the pdf

$$f(x) = \begin{cases} \frac{1}{100}, & 0 < x < 100 \\ 0, & \text{otherwise.} \end{cases} \Leftrightarrow X \sim U(0, 100)$$

$$\begin{aligned} E[X] &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_0^{100} x \frac{1}{100} dx = \frac{1}{100} \cdot \frac{1}{2} x^2 \Big|_0^{100} = 50 \end{aligned}$$

$$Var[X] = E[(X - E[X])^2] = \int_0^{100} (x - 50)^2 \frac{1}{100} dx = \frac{2500}{3}.$$

Mean and Variance for $U(a, b)$

Actually, for $X \sim U(a, b)$

$$E[X] = \frac{a+b}{2}, \quad Var[X] = \frac{(b-a)^2}{12},$$

They can be derived by

1. the definition

2. the mgf technique?

$$M(t) = \begin{cases} \frac{e^{tb}-e^{ta}}{t(b-a)}, & t \neq 0 \\ 1, & t = 0 \end{cases}$$

It does not work as usual and is skipped.

Example 4, page 99

Question

Let X be a continuous RV and have the pdf

$$f(x) = \begin{cases} xe^{-x}, & 0 < x < \infty \\ 0, & \text{otherwise} \end{cases}$$

$E[X]$ and $Var[X]$?

Example 4, page 99

Question

Let X be a continuous RV and have the pdf

$$f(x) = \begin{cases} xe^{-x}, & 0 < x < \infty \\ 0, & \text{otherwise} \end{cases}$$

$E[X]$ and $Var[X]$?

$$M(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx = \int_{0}^{\infty} xe^{-x} e^{tx} dx$$

$$= \int_{0}^{\infty} xe^{-(1-t)x} dx = \left[-\frac{xe^{-(1-t)x}}{1-t} - \frac{e^{-(1-t)x}}{(1-t)^2} \right] \Big|_{0}^{\infty}$$

Example 4, page 99

$$M(t) = \lim_{b\to\infty}\left[-\frac{be^{-(1-t)b}}{1-t} - \frac{e^{-(1-t)b}}{(1-t)^2}\right] + \frac{1}{(1-t)^2}$$

$$\underline{\underline{\text{when} t<1, i.e., 1-t>0}} \; \frac{1}{(1-t)^2}$$

$$M'(t) = 2\cdot\frac{1}{(1-t)^3} \Rightarrow M'(0) = 2$$

$$M''(t) = 6\cdot\frac{1}{(1-t)^4} \Rightarrow M''(0) = 6$$

$$E[X] = M'(0) = 2,$$

$$Var[X] = E[X^2] - (E[X])^2 = M''(0) - (M'(0))^2 = 2$$

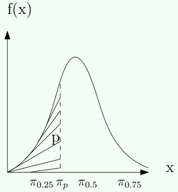
(100p)th percentile

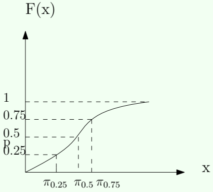
Definition

It is a number π_p such that the area under $f(x)$ to the left of π_p is p. That is

$$p = \int_{-\infty}^{\pi_p} f(x)dx = F(\pi_p)$$

The 50th percentile is called the median. The 25th and 75th percentiles are called the first and third quantiles, respectively. The median is also called the 2nd quantile.





Example 5

Let X be a continuous RV with the pdf

$$f(x) = \frac{3x^2}{4^3} e^{-(\frac{x}{4})^3}, \quad 0 < x < \infty$$

What is $\pi_{0.3}$?

Example 5

Let X be a continuous RV with the pdf

$$f(x) = \frac{3x^2}{4^3} e^{-(\frac{x}{4})^3}, \quad 0 < x < \infty$$

What is $\pi_{0.3}$?

$$F(x) = \int_{-\infty}^{x} f(y)dy = \begin{cases} 0, & -\infty < x < 0 \\ 1 - e^{(-\frac{x}{4})^3}, & 0 \leq x < \infty \end{cases}$$

$$F(\pi_{0.3}) = P(X \leq \pi_{0.3}) = 0.3$$
$$\Rightarrow 1 - e^{(-\frac{\pi_{0.3}}{4})^3} = 0.3, \quad \ln 0.7 = (-\frac{\pi_{0.3}}{4})^3$$
$$\Rightarrow \pi_{0.3} = -4(\ln 0.7)^{\frac{1}{3}} = 2.84$$

Section 3.2 Exponential, Gamma and Chi-square Distribution

Poisson Distribution

Now consider the approximate Poisson process (APP) with average number of occurrence λ in a unit interval.

Poisson distribution: let X describe the number of occurrences of some events in the unit interval with

$$f(x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, \cdots .$$

$$E[X] = \lambda, \quad Var[X] = \lambda$$

Number of occurrences in an interval with length T

For an interval with length T, which should be treated as a new "unit interval", the number of occurrences Y has a Poisson distribution with $E[Y] = \lambda T$ and thus its pmf is

$$f(y) = \frac{(\lambda T)^y e^{-\lambda T}}{y!}, \quad y = 0, 1, \cdots$$

Therefore, we have for any APP with average number of occurrence λ in a unit interval, the probability of having no occurrence in an interval with length T is

$$P(Y = 0) = e^{-\lambda T} = P(\text{no occurrence in the interval with length T})$$

Exponential distribution

1. Description: Consider an APP. We are interested in the waiting time until the first occurrence.
2. Define the waiting time by W. Then our goal is to derive the pdf of W.

Idea: $\begin{cases} \text{1.derive cdf of W, } F(w) \\ 2. f(w) = F'(w) \end{cases}$

$F(w) = P(W \leq w)$

Assume that the waiting time is nonnegative. Then,

$$F(w) = 0, \quad \text{for } w < 0.$$

For $w \geq 0$,

$$F(w) = P(W \leq w) = 1 - P(W > w)$$

Exponential distribution

where

$$P(W > w) = P(\text{no occurrences in [0,w]}) = e^{-\lambda w}$$

Therefore,

$$F(w) = 1 - e^{-\lambda w} \text{ for } w \geq 0$$

leading to

$$f(w) = F'(w) = \lambda e^{-\lambda w}, w \geq 0$$

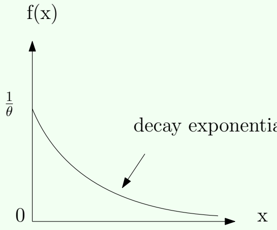
Exponential Distribution

Definition

A RV X has an exponential distribution if its pdf is

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad x \geq 0, \theta > 0$$

Accordingly, the waiting time until the first occurrence for an APP has an exponential distribution with $\theta = \frac{1}{\lambda}$ [λ: the average number of occurrences per unit time]



Mathematical expectations

3. mgf, mean and variance

$$M(t) = E[e^{tX}] = \int_0^{\infty} e^{tx} \cdot \frac{1}{\theta} e^{-\frac{x}{\theta}} dx$$

$$= \frac{1}{\theta} \frac{1}{(t - \frac{1}{\theta})} e^{(t-\frac{1}{\theta})x} \Big|_0^{\infty} = \frac{1}{1 - t\theta}, \quad t < \frac{1}{\theta}$$

$$\Rightarrow M'(t) = \frac{\theta}{(1-\theta t)^2}, \quad M''(t) = \frac{2\theta^2}{(1-\theta t)^3}$$

$$\Rightarrow M'(0) = \theta = E[X], \quad M''(0) = 2\theta^2 \Rightarrow Var[X] = \theta^2$$

Example 1, page 105

Question

Customers arrive in a shop according to APP at a mean rate of 20 per hour. What's the probability that the shopkeeper will have to wait more than 5 minutes for the arrival of the first customer?

Example 1, page 105

Question

Customers arrive in a shop according to APP at a mean rate of 20 per hour. What's the probability that the shopkeeper will have to wait more than 5 minutes for the arrival of the first customer?

Let X denote the waiting time in minute until the first customer arrives and note that $\lambda = \frac{1}{3}$ is the average number of customers per minute. Thus

$$\theta = \frac{1}{\lambda} = 3 \quad \text{and} \quad f(x) = \frac{1}{3} e^{-\frac{1}{3}x}, \quad x \geq 0$$

Hence

$$P(X > 5) = \int_5^\infty \frac{1}{3} e^{-\frac{1}{3}x} dx = e^{-\frac{5}{3}} \approx 0.18$$