Lecture Stat 302
Introduction to Probability - Slides 14

AD

March 2010

Continuous Random Variable

- **'Formal' definition**: We say that X is a (real-valued) continuous r.v. if there exists a *nonnegative* function $f : \mathbb{R} \rightarrow [0, \infty)$ such that for any set A of real numbers

$$P(X \in A) = \int_A f(x) \, dx.$$

- $f(x)$ is called the *probability density function* (pdf) of the r.v. X and the associated *(cumulative) distribution function* is

$$F(x) = \Pr(X \leq x) = \int_{-\infty}^{x} f(y) \, dy$$

so we have

$$f(x) = \frac{dF(x)}{dx}.$$

Example: Insurance Policy

- A group insurance policy covers the medical claims of the employees of a small company. The value, V, of the claims made in one year is described by

$$V = 100,000X$$

where X is a random variable with pdf

$$f(x) = \begin{cases} c(1-x)^4 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}.$$

- What is the conditional probability that V exceeds 40,000 given that V exceeds 10,000?

Example: Insurance Policy

- We are interested in

$$\begin{aligned} P\left(V>40,000 \mid V>10,000\right) &= \frac{P\left(V>40,000 \cap V>10,000\right)}{P\left(V>10,000\right)} \\ &= \frac{P\left(V>40,000\right)}{P\left(V>10,000\right)} \end{aligned}$$

where

$$P\left(V>v\right)=P\left(100,000X>v\right)=P\left(X>\frac{v}{100,000}\right)$$

- First we need to determine c using $\int_0^1 f\left(x\right)dx=1$; that is

$$\begin{aligned} \int_0^1 f\left(x\right)dx &= c\left[-\frac{\left(1-u\right)^5}{5}\right]_0^1=\frac{c}{5} \\ &\Rightarrow c=5. \end{aligned}$$

Example: Insurance Policy

- We need to compute the cdf $F_X(x)$of X which is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ c\left[-\frac{(1-u)^5}{5}\right]_0^x = 1-(1-x)^5 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

- So we are interested in

$$P(V > 40,000 | V > 10,000) = \frac{1-F_X(0.4)}{1-F_X(0.1)} = \frac{0.078}{0.590} = 0.132.$$

Example: Nuclear power plant

- Assume a nuclear power plant has three independent safety systems. These safety systems have lifetimes X_1, X_2, X_3 in years which are exponential r.v.s with respective parameters $\lambda_1 = 1$, $\lambda_2 = 0.5$ and $\lambda_3 = 0.1$. Since their installation five years ago, these systems have never been inspected. What is the proba that the nuclear power plant is currently being operated without any working safety system?
- The probability that the safety system i it is not working is

$$\begin{aligned} \Pr(X_i < 5) &= \lambda_i \int_0^5 \exp(-\lambda_i x)\, dx = 1 - \exp(-\lambda_i 5) \\ &= \begin{cases} 0.9933 & \text{if } i = 1 \\ 0.9179 & \text{if } i = 2 \\ 0.3935 & \text{if } i = 3 \end{cases} \end{aligned}$$

- Hence the probability that none of the system is working is simply

$$\Pr(X_1 < 5)\Pr(X_2 < 5)\Pr(X_3 < 5) = 0.3588$$

Expectation of Continuous Random Variables

- We define the expected valued of an r.v. X by

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

- More generally for any real-valued function $g : \mathbb{R} \rightarrow \mathbb{R}$ then

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

- **Uniform density.** We have for $c < d$ and $x \in [c, d]$

$$f(x) = \frac{1}{d - c}$$

then

$$E(X) = \int_{-\infty}^{\infty} x \ f(x) \, dx = \frac{1}{d-c} \int_{c}^{d} x dx = \frac{d^2 - c^2}{2(d-c)} = \frac{c+d}{2}.$$

- **Exponential density.** We have for $\lambda > 0$ and $x \geq 0$

$$f(x) = \lambda \exp(-\lambda x)$$

so

$$\begin{aligned} E(X) &= \lambda \int_0^\infty x \ \exp(-\lambda x)\, dx \\ &= \lambda \left[x \frac{\exp(-\lambda x)}{-\lambda} \right]_0^\infty - \lambda \int_0^\infty \frac{\exp(-\lambda x)}{-\lambda} dx \\ &= \frac{1}{\lambda}. \end{aligned}$$

- **Even density.** For $f(x) = f(-x)$, we have

$$\begin{aligned} & E(X) = \int_{-\infty}^0 x \ f(x)\, dx + \int_0^\infty x \ f(x)\, dx \\ & = \int_{-\infty}^0 x \ f(-x)\, dx + \int_0^\infty x \ f(x)\, dx \\ & = -\int_0^\infty u \ f(u)\, dx + \int_0^\infty x \ f(x)\, dx = 0 \end{aligned}$$

Expectation of Continuous Random Variables

- Consider the pdf

$$f(x) = \frac{1}{(x+1)^2} \text{ for } x \geq 0$$

then

$$\begin{aligned} E(X+1) &= \int_0^\infty \frac{x+1}{(x+1)^2} dx = \int_0^\infty \frac{1}{x+1} dx \\ &= \lim_{u \to \infty} [\log(x+1)]_0^u = \infty \end{aligned}$$

Hence we can conclude that $E(X)$ is infinite in this case.

- Distributions such that $E(X)$ is not finite are sometimes referred to as heavy-tails; they appear a lot in finance, actuarial science etc.

Example: Selling Printers

- The lifetime of a printer costing 200$ us exponentially distributed with mean 2 years. The manufacturer agrees to pay a full refund to a buyer if the printer fails during the first year following its purchase, and a one-half refund if it fails during the second year. If the manufacturer sells 100 printers, how much should it expect to pay in refunds?
- Let T denote a printer lifetime then

$$f(t) = \frac{1}{2} \exp\left(-\frac{t}{2}\right) \mathbf{1}_{(0,\infty)}(t)$$

so we have

$$\begin{aligned} P(T < 1) &= \int_0^1 f(t)\,dt = [\exp(-t/2)]_0^1 \\ &= 1 - \exp(-1/2) = 0.393, \\ P(1 < T < 2) &= \int_1^2 f(t)\,dt = [\exp(-t/2)]_1^2 \\ &= \exp(-1/2) - \exp(-1) = 0.239. \end{aligned}$$

Example: Selling Printers

- Let X_i denote the refund associated to the ith printer sold. Then for any $i = 1, ..., 100$

$$X_i = \begin{cases} 200 & \text{with proba } 0.393 \\ 100 & \text{with proba } 0.239 \\ 0 & \text{with proba } 0.368 \end{cases}$$

so we have

$$E(X_i) = 200 \times 0.393 + 100 \times 0.239 = 102.56.$$

- The expected refund associated to the 100 printers sold is thus

$$E\left(\sum_{i=1}^{100} X_i\right) = \sum_{i=1}^{100} E(X_i) = 100 \times 102.56 = 10,256.$$

Example: Failure Discovery

- A device that continuously measures and records seismic activity is placed in a remote region. The time, T, to failure of this device is exponentially distributed with mean 3 years. Since the device will not be monitored during its first two years of service, the time to discovery of its failure is $X = \max(T, 2)$. What is the expected value of X?
- We use the formula $E(g(T)) = \int g(t) f(t) \, dt$ for $f(t)$ an exponential of parameter $1/3$ and $g(t) = \max(t, 2)$ so

$$
\begin{aligned}
&E(X) = \int_0^\infty \max(t,2) \;\tfrac{1}{3} \exp\left(-\tfrac{t}{3}\right) dt \\
&= \int_0^2 \tfrac{2}{3} \exp\left(-\tfrac{t}{3}\right) dt + \int_2^\infty \tfrac{t}{3} \exp\left(-\tfrac{t}{3}\right) dt \\
&= \left[-2\exp\left(-\tfrac{t}{3}\right)\right]_0^2 - \left[t\exp\left(-\tfrac{t}{3}\right)\right]_2^\infty + \int_2^\infty \tfrac{1}{3}\exp\left(-\tfrac{t}{3}\right) dt \\
&= -2\exp\left(-\tfrac{2}{3}\right) + 2 + 2\exp\left(-\tfrac{2}{3}\right) - \left[3\exp\left(-\tfrac{t}{3}\right)\right]_2^\infty \\
&= 2 + 3\exp\left(-\tfrac{2}{3}\right) = 3.54
\end{aligned}
$$

Variance of Continuous Random Variables

- We define the variance as

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

- **Uniform density.** We have for $f(x) = \frac{1}{d-c}$ for $x \in [c, d]$ and $E(X) = \frac{c+d}{2}$. We also have

$$E(X^2) = \int_{-\infty}^{\infty} x^2 \ f(x)\, dx = \frac{1}{d-c}\int_{c}^{d} x^2 dx = \frac{d^3 - c^3}{3(d-c)} = \frac{c^2 + d^2 + cd}{3}$$

so

$$Var[X] = \frac{c^2 + d^2 + cd}{3} - \frac{(c+d)^2}{4} = \frac{(d-c)^2}{12}$$

Example: Repair Cost and Insurance Payement

- The owner of an automobile insures it against damage by purchasing an insurance policy with a deductible of 250$. In the event that the automobile is damaged, repair costs can be modeled by a uniform random variables on the interval $(0, 1500)$. Determine the standard deviation of the insurance payement in the event that the automobile is damaged.
- Let X be the repair cost and Y the insurance payement then

$$Y = \begin{cases} 0 & \text{if } X < 250 \\ X - 250 & \text{if } X \geq 250 \end{cases}$$

and we want to compute $\sqrt{Var(Y)}$.

Example: Repair Cost and Insurance Payement

- We have

$$E(Y) = \int_{250}^{1500} \frac{1}{1500}(x-250)\,dx = \frac{1}{3000}\left[(x-250)^2\right]_{250}^{1500} = 521,$$
$$E\left(Y^2\right) = \int_{250}^{1500} \frac{1}{1500}(x-250)^2\,dx = \frac{1}{4500}\left[(x-250)^3\right]_{250}^{1500} = 434,028.$$

- Finally, we obtain

$$\begin{aligned} Var(X) &= E\left(Y^2\right) - E(Y)^2 = 434,028 - 521^2, \\ \sqrt{Var(Y)} &= 403. \end{aligned}$$

- **Exponential density.** We have $f(x) = \lambda \exp(-\lambda x)$ for $\lambda > 0$ and $x \geq 0$ and $E(X) = \frac{1}{\lambda}$. We have

$$
\begin{aligned}
E\left(X^2\right) &= \lambda \int_0^\infty x^2 \ \exp(-\lambda x)\, dx \\
&= \lambda \left[x^2 \frac{\exp(-\lambda x)}{-\lambda} \right]_0^\infty - \lambda \int_0^\infty 2x \frac{\exp(-\lambda x)}{-\lambda} dx \\
&= \frac{2}{\lambda} E(X) = \frac{2}{\lambda^2}
\end{aligned}
$$

so

$$Var(X) = E\left(X^2\right) - E(X)^2 = \frac{1}{\lambda^2}.$$

Median of Continuous Random Variables

- The median of a continuous r.v. X of pdf $f(x)$ is the number m such that
$$\int_{-\infty}^{m} f(x)\,dx = \int_{m}^{\infty} f(x)\,dx = \frac{1}{2};$$
that is the number m such that
$$\Pr(X \leq m) = P(X \geq m) = \frac{1}{2}.$$
- For example, assume we look at a population of people. Let X be the salary of a randomly chosen person from this population of pdf $f(x)$, and let m be the median salary of the population. This means that half the population earns less than m dollars and half earns more than m dollars.
- **Uniform density.** For $c < d$, we have $f(x) = \frac{1}{d-c}$ and the median is $m = \frac{c+d}{2}$; i.e. in this case the median and $E(X)$ are similar.

Median of Continuous Random Variables

- **Exponential density.** We have $f(x) = \lambda \exp(-\lambda x)$ for $\lambda > 0$ and $x \geq 0$.
- The median corresponds to the value

$$\int_0^m \lambda \exp(-\lambda x)\, dx = \int_m^\infty \lambda \exp(-\lambda x)\, dx = \frac{1}{2}.$$

- We have

$$\begin{aligned} \int_m^\infty \lambda \exp(-\lambda x)\, dx &= [\exp(-\lambda x)]_m^\infty \\ &= \exp(-\lambda m) \end{aligned}$$

and

$$\exp(-\lambda m) = \frac{1}{2} \Leftrightarrow m = \frac{\log 2}{\lambda}$$

- In this case, the median and $E(X)$ are different.