# Lecture Stat 302
# Introduction to Probability - Slides 13

AD

March 2010

# Continuous Random Variables

- For the time being, we have only considered discrete random variables (r.v.) - set of possible values is finite or countable - such as the number of arrivals in a given time instants, the number of successful trials, the number of items with a given characteristic.
- In many practical applications, we have to deal with random variables whose set of possible values is uncountable.
- **Example: Waiting for a bus**. The time (in minutes) which elapses between arriving at a bus stop and a bus arriving can be modelled as a r.v. $X$ taking values in $[0, \infty)$.
- **Example: Share price.** The values of one share of a specific stock at some given future time can be modelled as a r.v. $X$ taking values in $[0, \infty)$.
- **Example: Weight**. The weight of a randomly chosen individual can be modelled as a r.v. $X$ taking values in $[0, \infty)$.
- **Example: Temperature.** The temperature in Celsius at a given time can be modelled as a r.v. $X$ taking values in $[-273.15, \infty)$.

# Continuous Random Variable

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

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

- We also have

$$P(X \in \mathbb{R}) = 1 = \int_{\mathbb{R}} f(x)\, dx$$

# Continuous Random Variable

- Assume we have a set $A = [a, b]$ then

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

- $P(X \in [a, b])$ corresponds to the area under the curve $f(x)$ between $a$ and $b$.
- Consider a small interval $[x, x + \Delta x]$, then we have

$$P(x \leq X \leq x + \Delta x) = \int_x^{x+\Delta x} f(y)\,dy \approx \; f(x) \; \Delta x$$

so $P(x \leq X \leq x + \Delta x)$ is *proportional* to $f(x)$ for $\Delta x$ small.

# Examples of pdf

- **Uniform density.** Consider a r.v. $X$ with pdf given for $c < d$

$$f(x) = \begin{cases} \frac{1}{d-c} & \text{if } x \in [c, d] \\ 0 & \text{otherwise} \end{cases}$$

- **Exponential density.** Consider a r.v. $X$ with pdf given for $\lambda > 0$ by

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } x \in [0, \infty) \\ 0 & \text{otherwise} \end{cases}$$

- **Beta density.** Consider a r.v. $X$ with pdf given for $\alpha, \beta > 0$ by

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{if } x \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

# Important Remarks

- $f(x)$ is NOT a probability, e.g. $f(1)$ is not the probability that $X = 1$, it is a probability density.
- As $f(x)$ is NOT a probability, it is NOT surprising to have $f(x) > 1$ for some values of $x \in \mathbb{R}$.
- For a continuous r.v., the probability of any single value is zero; i.e. for any $x \in \mathbb{R}$ we have $P(X = x) = 0$. For example if you model a waiting time $X$ by an exponential density, then the proba of waiting 5,123123123123... or exactly 5.0 minutes will be zero!
- $P(X = x) = 0$ for any $x$ implies that $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$.
- Keep track of the "range" of a density, such as the interval $[0, 1]$ or $[0, \infty)$ when doing your calculations.

# Toy example

- Assume you are told that the pdf of a r.v. $X$ satisfies
$$f(x) = \begin{cases} Cx & \text{if } x \in [0,1] \\ Cx^2 & \text{if } x \in [-1,0] \\ 0 & \text{otherwise.} \end{cases}$$
where $C$ is an unknown constant. What is the probability that $X < 0.5$.
- We first determine the value of $C$ using the fact that $\int_{\mathbb{R}} f(x)\,dx = 1$;
$$\begin{aligned} \int_{\mathbb{R}} f(x)\,dx &= C\int_{-1}^{0} x^2 dx + C\int_0^1 xdx \\ &= \frac{C}{3} + \frac{C}{2} = \frac{5C}{6} = 1 \Rightarrow C = \frac{6}{5}. \end{aligned}$$
- The probability that $X < 0.5$ is
$$\Pr(X < 0.5) = \frac{6}{5}\left[\int_{-1}^{0} x^2 dx + \int_0^{0.5} xdx\right] = \frac{6}{5}\left(\frac{1}{3} + \frac{1}{8}\right)$$

# Another toy example

- Consider the following pdf

$$f(x) = \begin{cases} \frac{c}{(x+1)^2} & \text{for } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

  What is the value of $c$?

- We are going to use $\int_{\mathbb{R}} f(x)\, dx = 1$

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\, dx &= \int_0^{\infty} \frac{c}{(x+1)^2} dx = \left[ -\frac{c}{x+1} \right]_0^{\infty} \\ &= c \end{aligned}$$

  so $c = 1$.

# Example: Age distribution

- Assume (completely wrongly) that the density $f(x)$ of people's ages in years in Canada is constant between $[0, 50]$ and then decreases linearly to zero between $[50, 100]$. (Note: this model does not authorize people over one hundred years).
- What is the expression of $f(x)$?
- What is the probability that a random chosen person's age is between 40 and 50?

# Example: Age distribution

- We have

$$f(x) = \begin{cases} \alpha & \text{for } x \in [0, 50] \\ \alpha - \beta(x - 50) & \text{for } x \in [50, 100] \\ 0 & \text{otherwise} \end{cases}$$

  where $\alpha - \beta(100 - 50) = \alpha - 50\beta = 0$ and we can solve $\int f(x)\, dx = 1$ to obtain $\alpha$.

- We can also make a simple graph to show directly that

$$\begin{aligned} \int f(x)\, dx &= 50\alpha + 50\alpha/2 = 75\alpha = 1 \\ &\Rightarrow \alpha = \frac{1}{75}. \end{aligned}$$

- Hence we have

$$\Pr(40 \leq X \leq 50) = \int_{40}^{50} f(x)\, dx = \alpha(50 - 40) \approx 0.13$$

# Example: Waiting times

- We have previously discussed the Poisson distribution. If the average rate of events is $\lambda$ in a given time interval then it can be shown that the interarrival time between two events is a continuous r.v. with a so-called exponential density

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } x \in [0, \infty) \\ 0 & \text{otherwise} \end{cases} .$$

- Assume you receive an average of 2 phones calls per hour. You make the Poisson assumption so the time between two events follows an exponential density. Assume you have just hung up the phone, (a) what is the probability that you are going to have to wait more than one hour to receive the next call? (b) what is the probability that you are going to receive a phone call in the next ten minutes? (c) what is the probability you are going to have to wait more than 10 minutes but less than one hour?

# Example: Waiting times

- Let $X$ the time in hour you have to wait, this is a r.v. with an exponential pdf of parameter $\lambda = 2$.
- (a) We first want to compute $P(X > 1)$. We have

$$P(X > 1) = \lambda \int_1^\infty \exp(-\lambda x)\, dx = \exp(-\lambda) \approx 0.13.$$

- (b) We want to compute $P\left(X < \frac{10}{60}\right)$ which is given by

$$P\left(X < \frac{10}{60}\right) = \lambda \int_0^{1/6} \exp(-\lambda x)\, dx = 1 - \exp(-\lambda/6) \approx 0.28.$$

- (c) We want to compute $P\left(\frac{10}{60} < X < 1\right)$

$$\begin{aligned} P\left(\frac{10}{60} < X < 1\right) &= P(X < 1) - P\left(X < \frac{10}{60}\right) \\ &= (1 - P(X > 1)) - P\left(X < \frac{10}{60}\right) \\ &\approx 0.58. \end{aligned}$$