Stats 426 Lecture 4

Thomas Cohn

2/1/21

Continuous Random Variables and Probability Distributions

Recall: A random variable X is continuous if its set of possible values contains an entire interval of numbers. The distribution of X is characterized by a pdf $f(x)$ s.t. $\forall a, b \in \mathbb{R}$ with $a < b$, $P(a < X < b) = \int_a^b f(x)\,dx$. Also, $f(x) \geq 0\ \forall x \in \mathbb{R}$, f is piecewise continuous, and $\int_{\mathbb{R}} f = 1$.

Defn: Suppose X is a continuous random variable with density function f. Then the cumulative distribution function (or cdf) of X is

$$F(x) \stackrel{\text{def}}{=} P(X \leq x) = \int_{-\infty}^{x} f(u)\,du$$

Observe: $\lim_{s \to -\infty} F(s) = 0$, $\lim_{s \to \infty} F(s) = 1$.

Some facts about X, an arbitrary continuous random variable:

- $P(X = c) = 0$, $\forall c \in \mathbb{R}$
- $P(a \leq X \leq b) = F(b) - F(a)$
- $P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b)$
- $\forall x$ where $F'(x)$ exists, $\frac{d}{dx} F(x) = f(x)$.

Quantiles

Defn: If a cdf F is strictly increasing, for $p \in (0,1)$, $x = F^{-1}(p)$ is called the pth quantile.

Defn: The 0.5 quantile is the median.
The 0.25 quantile is the lower quantile.
The 0.75 quantile is the upper quantile.

Defn: X follows an exponential distribution with parameter λ ($X \sim \text{Exp}(\lambda)$) if

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

Note: We must have $\lambda > 0$.

Exp is often used as the distribution of lifetimes, or times between the occurrence of successive events.

Suppose $X \sim \text{Exp}(\lambda)$, and $x, x_0 > 0$. Then $P(X \geq x + x_0 \mid X \geq x_0) = P(X \geq x)$.

Proof:

$$P(X \geq x + x_0 \mid X \geq x_0) = \frac{P(X \geq x + x_0, X \geq x_0)}{P(X \geq x_0)} = \frac{P(X \geq x + x_0)}{P(X \geq x_0)} = \frac{1 - (1 - e^{-\lambda(x + x_0)})}{1 - (1 - e^{-\lambda x_0})} = e^{-\lambda x} = P(X \geq x)$$

$\square$

Ex: The lifetime of a component is exponentially-distributed, with parameter $\lambda = 3$. If it has already worked for 10 hours, what is the probability that it works for 4 more hours?

$$P(X \geq 10 + 4 \mid X \geq 10) = P(X \geq 4) = e^{-3 \cdot 4} = e^{-12}$$

Defn: X follows a gamma distribution with shape parameter $\alpha > 0$ and scale parameter $\lambda > 0$ ($X \sim \gamma(\lambda, \alpha)$) if

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & x \geq 0 \\ 0 & x < 0 \end{cases}$$

where the gamma function Γ satisfies

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx$$

Some properties of the gamma function:

- $\forall \alpha > 1$, $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$
- $\forall n \in \mathbb{N}$, $\Gamma(n) = (n-1)!$
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$

Some properties of the gamma distribution:

- Exp is a special case with $\alpha = 1$.
- The sum of i.i.d. exponential random variables follows a gamma distribution. If $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$, then

$$Y = \sum_{i=1}^n X_i \sim \Gamma(\lambda, n)$$

- The sum of i.i.d. gamma random variables follows a gamma distribution. If $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} \Gamma(\lambda, \alpha)$, then

$$Y = \sum_{i=1}^n X_i \sim \Gamma(\lambda, n\alpha)$$

Defn: X follows a beta distribution with parameters a, b ($X \sim \text{Beta}(a, b)$) if, for $0 \leq x \leq 1$,

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$$

Observe: When $a = b = 1$, $X \sim \text{Unif}[0, 1]$.

Defn: The normal distribution, with $\mu = 0$ and $\sigma = 1$, is called the standard normal distribution. $Z \sim \text{N}(0, 1)$ is called a standard normal random variable, with

$$f(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad F(z) = \Phi(z) = P(Z \leq z)$$

We can standardize any normal random variable: $X \sim \text{N}(\mu, \sigma^2) \Leftrightarrow Z = \frac{X-\mu}{\sigma} \sim \text{N}(0, 1)$. Now, $P(X \leq x) = \Phi(\frac{x-\mu}{\sigma})$.

Quantiles of the Normal Distribution

Notation: z_α is the $100(1-\alpha)$ percentile / $1-\alpha$ quantile of the standard normal distribution. Say $X \sim \text{N}(\mu, \sigma^2)$. Let η_p be the p quantile of X. By standardizing, $P(X \leq \eta_p) = \Phi(\frac{\eta_p - \mu}{\sigma})$. We then solve $\Phi(\frac{\eta_p-\mu}{\sigma}) = p$, and get $z_{1-p} = \frac{\eta_p - \mu}{\sigma}$, i.e., $\eta_p = \mu + \sigma z_{1-p}$.

Say X is a random variable with pdf f_X and cdf F_X. Let $Y = g(X)$. We want to find $f_Y = f_{g(X)}$.

Approach 1: The direct approach – work through the cdf. Derive F_Y and differentiate.

Ex: $X \sim \text{N}(0, 1)$, $Y = X^2$. Then $f_Y(y) = y^{-1/2}\phi(\sqrt{y})$ for $y \geq 0$, and 0 otherwise.

Proof: For $y \geq 0$,

$$F_Y(y) = P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y})$$

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy}(\Phi(\sqrt{y}) - \Phi(-\sqrt{y})) = \frac{1}{2}\phi(y) y^{-1/2} - (-1 \cdot \tfrac{1}{2})\phi(y) y^{-1/2} = \phi(y) y^{-1/2}$$

□

Ex: If $X \sim \mathrm{N}(\mu, \sigma^2)$, $Y = aX + b$, then $Y \sim \mathrm{N}(a\mu + b, a^2\sigma^2)$.

Approach 2: Chain rule.

Thm: (Chain Rule) Suppose f, g are differentiable, and $h = g \circ f$. Then $h' = (g' \circ f)f'$.

Now suppose g is differentiable, and monotonic on an interal I, with $f_X(x) = 0$ outside of I. Then

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right| = f_X(g^{-1}(y)) \left| \frac{1}{g'(g^{-1}(y))} \right|$$

for all y s.t. $\exists x \in I$ s.t. $y = g(x)$. Otherwise, $f_Y(y) = 0$.

Defn: The <u>joint probability mass function</u> of a pair of discrete random variables X, Y is $p(x, y) = P(X = x \wedge Y = y)$.

The joint pmf must satisfy

- $p(x, y) \geq 0$, $\forall (x, y)$
- $\sum_x \sum_y p(x, y) = 1$

Defn: The <u>marginal probability mass function</u> of X, given a joint probability mass function p, is $p_X(x) = \sum_y p(x, y)$.

Defn: X and Y are <u>independent</u> if $\forall (x, y)$, $p(x, y) = p_X(x)p_Y(y)$.

Defn: The <u>cumulative distribution function</u> of X and Y is

$$F(x, y) = P(X \leq x, Y \leq y) = \sum_{\substack{x_i \leq x \\ y_i \leq y}} p(x_i, y_i)$$

For $X_1, \ldots, X_m$, the pdf is $p(x_1, \ldots, x_m) = P(X_1 = x_1, \ldots, X_m = x_m)$.
The marginal pdf of X_k is

$$p_{X_k}(x) = \sum_{\substack{x_1, \ldots, x_{k-1}, \\ x_{k+1}, \ldots, x_m}} p(x_1, \ldots, x_{k-1}, x, x_{k+1}, \ldots, x_m)$$

The marginal pdf of X_k and X_ℓ is (WOLOG $k < \ell$)

$$p_{X_k, X_\ell}(s, t) = P(X_k = s, X_\ell = t) = \sum_{\substack{x_1, \ldots x_{k-1}, \\ x_{k+1}, \ldots, x_{\ell-1}, \\ x_{\ell+1}, \ldots, x_m}} p(x_1, \ldots, x_{k-1}, s, x_{k+1}, \ldots, x_{\ell-1}, t, x_{\ell+1}, \ldots, x_m)$$

Ex: A fair coin is tossed 3 times. X is 1 if the first toss is heads, and 0 if it's tails. Y is the total number of heads. Then the sample space is

$$\{(H, H, H), (H, H, T), (H, T, H), (H, T, T), (T, H, H), (T, H, T), (T, T, H), (T, T, T)\}$$

The pdf of (X, Y) is

		Y			
	$p(x, y)$	0	1	2	3
X	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	0
	1	0	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$

$P(X = 0, Y = 2) = \frac{1}{8}$
$P(Y = 2) = P(X = 0, Y = 2) + P(X = 1, Y = 2) = \frac{3}{8}$
$F_{X,Y}(0, 2) = P(X \leq 0, Y \leq 2) = \frac{1}{8} + \frac{2}{8} + \frac{1}{8} = \frac{1}{2}$
The marginal cdf of Y is

Y	0	1	2	3
pdf	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
cdf	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{7}{8}$	1