STA 623 – Fall 2009 – Dr. Charnigo

Section 1.6: Density and Mass Functions

Probability mass function. Let X be a discrete random variable. We define the probability mass function of X as $f(x) := P(X = x)$ for any $x \in \mathbb{R}$. Note that (lower case) x is a placeholder for 2 or 8 or some other number, so that (for example) $f(2) = P(X = 2)$ and $f(8) = P(X = 8)$.

A probability mass function must satisfy $f(x) \geq 0$ for all $x \in \mathbb{R}$ and $\sum_{x \in \mathbb{R}: f(x) > 0} f(x) = 1$. Conversely, any function $f(x)$ with these properties may be interpreted as a probability mass function.

For a discrete random variable X with probability mass function $f(x)$ and cumulative distribution function $F(x)$, we have $P(X \in B) = \sum_{x \in B: f(x) > 0} f(x)$ for any set $B \in \mathcal{B}^1$, the sigma field on $\mathbb{R}$ generated by its open subintervals. In particular,

$$\begin{aligned} & P(a \leq X \leq b) = \sum_{x \in [a,b]: f(x) > 0} f(x) \\ \geq \ & P(a < X \leq b) = \sum_{x \in (a,b]: f(x) > 0} f(x) = F(b) - F(a) \\ \geq \ & P(a < X < b) = \sum_{x \in (a,b): f(x) > 0} f(x) \end{aligned}$$

for any $a, b \in \mathbb{R}$ with $a < b$. The first "$\geq$" above is "$>$" if $f(a) > 0$ and "$=$" otherwise. The second "$\geq$" above is "$>$" if $f(b) > 0$ and "$=$" otherwise.

Example (probability mass function). Let λ be a positive number and put $f(x) := 1_{\{x \in \{0,1,2,\ldots\}\}} C(\lambda) \lambda^x / x!$. How can $C(\lambda)$ be chosen so that $f(x)$ is a probability mass function?

Let X be a random variable with this probability mass function, which we may write in shorthand as $X \sim f(x)$. What is $P(X > 1)$? What is $P(X \text{ even})$?

Probability density function. Let X be a continuous random variable with cumulative distribution function $F(x)$. Suppose that there exists a function $f(x)$ such that $F(x) = \int_{-\infty}^{x} f(t)\ dt$. Then we refer to $f(x)$ as a probability density function of X. Since altering $f(x)$ at finitely many points has no impact on its integration, a probability density function is not unique.

A probability density function must satisfy $\int_{-\infty}^{\infty} f(t)\ dt = 1$ and cannot be negative over any interval of nonzero length. We may as well assume, as is routinely done, that $f(x) \geq 0$ for all $x \in \mathbb{R}$. Conversely, any function $f(x)$ with these properties may be interpreted as a probability density function.

If $f(x)$ is continuous, then $f(x) = \frac{d}{dx}F(x)$.

For a continuous random variable X with probability density function $f(x)$ and cumulative distribution function $F(x)$, we have

$$P(a < X < b) = \int_{a}^{b} f(x)\ dx$$

for any $a, b \in \mathbb{R}$ with $a < b$. The above equality also holds with $a = -\infty$ and/or $b = \infty$. For finite a and b, we have the additional equalities

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

Example (probability density function). Let λ be a positive number and put $f(x) := 1_{\{x>0\}} C(\lambda) \exp[-\lambda x]$. How can $C(\lambda)$ be chosen so that $f(x)$ is a probability density function?

Let X be a random variable with this probability density function, which we may write in shorthand as $X \sim f(x)$. What is $P(X > 1)$? What is $F(x)$? Note that $f(x)$ is discontinuous only at 0 and that $f(x) = \frac{d}{dx}F(x)$ for all $x \neq 0$.

Additional practice. Let X be a random variable with probability density function $f(x) := 1_{\{x>0\}}(\log 2)\exp[-(\log 2)x]$. Show that, for any positive reals a, b, we have

$$P(X \geq a + b \mid X \geq a) = P(X \geq b \mid X \geq 0).$$

If X is used to model the lifetime of a light bulb in months, then the preceding equality says that a light bulb that has been in use for a months is just as likely to last another b months as is a fresh light bulb.

Let $Y := \lceil X \rceil$, read "the ceiling of X", which means the least integer greater than or equal to X. What is the probability mass function of Y? Show that, for any positive integers a, b, we have

$$P(Y > a + b \mid Y > a) = P(Y > b \mid Y > 0).$$

If Y is used to model the number of coin flips required to obtain one's first "Heads", then the preceding equality says that a person who has already started flipping a coin and obtained "Tails" on each of the first a attempts is just as likely to get all "Tails" in the next b attempts as is another person who has just started flipping a coin.