Stats 426 Lecture 3

Thomas Cohn

1/27/21

Defn: A permutation is an ordered arrangement of a set of objects. A combination is an unordered set of objects.

Defn: $\binom{n}{k}$ is a binomial coefficient.

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

Defn: A random variable is a numerically-valued function on a sample space Ω. Random variables are normally denoted with capital letters.

Ex: Tossing a coin. $\Omega = \{H, T\}$. $X(H) = 0$, $X(T) = 1$.

Random variables are typically discrete or continuous (although a mixture is possible), determined by their corresponding sample space.

Notation: The event $X = x \rightsquigarrow \{\omega \in \Omega \mid X(\omega) = x\}$.
The event $a \leq X \leq B \rightsquigarrow \{\omega \in \Omega \mid a \leq X(\omega) \leq b\}$.

Defn: The probability distribution of a discrete random variable is a list of distinct values x_i of X_i, together with their associated probabilities, denoted

$$p(x) = P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\})$$

Another name is the probability mass function or probability density function. It's commonly abbreviated pmf or pdf.

Defn: The cumulative distribution function (or cdf) of a discrete random variable is the function

$$F(x) = P(X \leq x) = \sum_{i : x_i \leq x} p(x_i)$$

for $-\infty < x < \infty$.

Ex: Tossing a coin.

s	H	T
X(s)	0	1

- $p(0) = P(X = 0) = 0.5$
- $p(1) = P(X = 1) = 0.5$
- $F(0) = P(X \leq 0) = P(X = 0) = 0.5$
- $F(0.5) = P(X \leq 0.5) = P(X = 0) = 0.5$
- $F(1) = P(X \leq 1) = P(X = 0) + P(X = 1) = 1$

Suppose we don't know if the coin is fair. Instead, we have some $0 < \alpha < 1$ where $p(0) = 1 - \alpha$, $p(1) = \alpha$. We write

$$p(x; \alpha) = \begin{cases} 1 - \alpha & x = 0 \\ \alpha & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Here, α is called a parameter of the distribution, and the collection of distributions $\{p(x; \alpha) \mid \alpha \in \mathcal{A}\}$ is called a family of distributions. $\mathcal{A}$ is the parameter space.

Defn: A Bernoulli random variable is a random variable whose only possible values are 0 and 1. We say $X \sim \mathrm{Ber}(p)$, and $P(X=1)=p$, $P(X=0)=1-p$, for $p \in [0,1]$.

Defn: A Binomial random variable is the sum of independent and identically distributed (i.i.d.) Bernoulli random variables. Given $Z_1, \dots, Z_n \sim \mathrm{Ber}(p)$, $X = \sum_{i=1}^{n} Z_i$, we say $X \sim \mathrm{Bin}(n;p)$. The pmf for X is

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Defn: A Geometric random variable is the number of i.i.d. Bernoulli trials before a success. If $Z_i \sim \mathrm{Ber}(p)$, we say $X \sim \mathrm{Geom}(p)$, and its pmf is

$$P(X=k) = (1-p)^{k-1} p \qquad k = 1, 2, \dots$$

Defn: A Negative Binomial random variable is the number of i.i.d. Bernoulli trials before r successes (for fixed r). If $Z_i \sim \mathrm{Ber}(p)$, then the pmf of X is

$$P(X=k) = \binom{k-1}{r-1} p^r (1-p)^{k-r} \qquad k \geq r$$

Defn: A Poisson random variable counts the number of events in a specified interval or region. λ is the parameter representing the rate. For $X \sim \mathrm{Poisson}(\lambda)$, the pmf is

$$P(X=x) = \frac{e^{-\lambda}\lambda^k}{x!}$$

We derive the Poisson distribution as the limit of a binomial distribution. Say $X \sim \mathrm{Bin}(n,p)$, and $p \to 0$, $n \to \infty$, and $np \approx \lambda$. Then $p(x) \to \frac{e^{-\lambda}\lambda^x}{x!}$:

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^n = \frac{\lambda^x}{x!} \frac{n!}{(n-x)! n^x} \left(1 - \frac{\lambda}{n}\right)^{-x} \left(1 - \frac{\lambda}{n}\right)^n$$

Now, we take the limit

$$\lim_{n\to\infty} P(X=x) = \frac{\lambda^x}{x!} \lim_{n\to\infty} \underbrace{\frac{n(n-1)\cdots(n-x+1)}{n^x}}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-x}}_{\to 1} \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} = \frac{\lambda^x e^{-\lambda}}{x!}$$

□