Topic 9

Examples of Mass Functions and Densities

Discrete Random Variables

Outline

Introduction

Write

$$f_X(x|\theta) = P_\theta\{X = x\}$$

for the mass function of the given family of discrete random variables depending on the parameter θ. We will

- use the expression $Family(\theta)$ as shorthand for this family
- followed by the R command `family` and
- the state space S.

Bernoulli Random Variables

$Ber(p)$ on $S = \{0, 1\}$

$$f_X(x|p) = \left\{ \begin{array}{ll} 0 & \text{with probability } (1-p), \\ 1 & \text{with probability } p, \end{array} \right\} = p^x(1-p)^{1-x}.$$

This is the simplest random variable, taking on only two values, namely, 0 and 1.

Think of it as the outcome of a Bernoulli trial, i.e., a single toss of an unfair coin that turns up heads with probability p.

Binomial Random Variables

$Bin(n,p)$ (R command `binom`) on $S = \{0, 1, \ldots, n\}$

$$f_X(x|p) = \binom{n}{x} p^x (1-p)^{n-x}.$$

The binomial distribution arises from computing the probability of x successes in n Bernoulli trials.

Considered in this way, the family $Ber(p)$ is also $Bin(1,p)$.

Exercise. Enter in R `plot(c(0:12),dbinom(c(0:12),12,p,),type="h")` for `p=0.25,0.5,0.75`. Describe what you see.

Negative Binomial Random Variables

Negbin(n, p) (R command `nbinom`) on $S = \mathbb{N}$

$$f_X(x|p) = \binom{n+x-1}{x} p^n (1-p)^x.$$

This random variable is the number of *failed* Bernoulli trials before the n-th success.
To find the mass function,

- For the outcome $\{X = x\}$, the n-th success must occur on the $n+x$-th trial. So,
- we must have $n-1$ successes and x failures in first $n+x-1$ Bernoulli trials
- followed by success on the last trial.

Negative Binomial Random Variables

The first $n + x - 1$ trials and the last trial are independent and so the probabilities multiply.

$$\begin{aligned} P_p\{X = x\} &= P_p\{n-1 \text{ successes in } n+x-1 \text{ trials, success in the } n-x\text{-th trial}\} \\ &= P_p\{n-1 \text{ successes in } n+x-1 \text{ trials}\} P_p\{\text{success in the } n-x\text{-th trial}\} \\ &= \binom{n+x-1}{n-1} p^{n-1}(1-p)^x \cdot p = \binom{n+x-1}{x} p^n (1-p)^x \end{aligned}$$

The first factor is computed from the binomial distribution, the second from the Bernoulli distribution. Note the use of the identity

$$\binom{m}{k} = \binom{m}{m-k}$$

in giving the final formula.

Poisson Random Variables

$Pois(\lambda)$ (R command `pois`) on $S = \mathbb{N}$

$$f_X(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}.$$

The Poisson distribution approximates of the binomial distribution for n large, p small, but the product $\lambda = np$ is moderate.

Examples.

- bacterial colonies
  - n is the number of bacteria and p is the probability of a mutation,
- recombination events during meiosis
  - n is the number of nucleotides on a chromosome and p is the probability of a recombination event occurring at a particular nucleotide.

The product λ is the mean number of events.

Poisson Random Variables

The approximation is based on the limit

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}.$$

We compute binomial probabilities, replace p by λ/n and take a limit as $n \to \infty$.

In this computation, we use the fact that for a fixed value of x,

$$\frac{(n)_x}{n^x} \to 1 \quad \text{and} \quad \left(1-\frac{\lambda}{n}\right)^{-x} \to 1 \quad \text{as } n \to \infty$$

Poisson Random Variables

$$P\{X=0\} = \binom{n}{0}p^0(1-p)^n = \left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda}$$

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

$$P\{X=2\} = \binom{n}{2}p^2(1-p)^{n-2} = \frac{n(n-1)}{2}\left(\frac{\lambda}{n}\right)^2\left(1-\frac{\lambda}{n}\right)^{n-2} = \frac{n(n-1)}{n^2}\frac{\lambda^2}{2}\left(1-\frac{\lambda}{n}\right)^{n-2} \approx \frac{\lambda^2}{2}e^{-\lambda}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

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

Exercise. Explain why

$$\sum_{x=0}^{\infty} f_X(x) = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!}e^{-\lambda} = 1.$$

Hypergeometric Random Variables

$Hyper(m, n, k)$ (R command `hyper`) on $S = \{\max\{0, k - n\}, \ldots, \min\{m, k\}\}$

$$f_X(x|m, n, k) = \frac{\binom{m}{x}\binom{n}{k-x}}{\binom{m+n}{k}}$$

- Begin with an urn holding m white balls and n black balls.
- Remove k at random and
- let the random variable X denote the number of white balls.

Hypergeometric Random Variables

We consider equally likely outcomes to determine $f_X(x|m,n,k) = P_{m,n,k}\{X = x\}$.

- The total number of possible outcomes, $\#(\Omega)$, namely, the number of ways to choose k balls out of an urn containing $m+n$ balls.

$$\binom{m+n}{k}.$$

  This will be the denominator for the probability.

- For the numerator, the outcomes that result in x white balls from the total of m, we must also choose $k-x$ black balls from the total of n. By the multiplication property, the number of ways $\#(A_x)$ to accomplish this is product.

$$\binom{m}{x}\binom{n}{k-x}.$$

Finally, $f_X(x|m,n,k) = \#(A_x)/\#(\Omega)$.