Chapter 3

# 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 $\theta$. We will

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

# Discrete Uniform Random Variables

$Unif(a,b)$(R command `sample`) on $S = \{a, a+1, \ldots, b\}$,

$$f_X(x|a,b) = \frac{1}{b-a+1}$$

As we have seen, for $X$, a $Unif(1,n)$ random variable,

$$EX = \frac{n+1}{2} \qquad \mathrm{Var}(X) = \frac{n^2-1}{12}.$$

The probability generating function

$$\rho_X(z) = \frac{1}{n}(z + z^2 + \cdots + z^n) = \frac{1}{n} \cdot \frac{z(1-z^n)}{1-z}.$$

# 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 $p$.

$$EX = p, \qquad \mathrm{Var}(X) = p(1-p), \qquad \rho_X(t) = (1-p) + pz.$$

# Binomial Random Variables

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

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

Previous computations have shown us that

$$EX = np, \qquad \mathrm{Var}(X) = np(1-p), \qquad \rho_X(t) = ((1-p) + pz)^n.$$

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

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

# Geometric Random Variables

$Geom(p)$ (R command `geom`) on $S = \mathbb{N}$

$$f_X(x|p) = p(1-p)^x.$$

The geometric distribution arises from computing the probability of the number of *failed* Bernoulli trials before the first success.

For this random variable, we have

$$EX = \frac{1-p}{p}, \quad \mathrm{Var}(X) = \frac{1-p}{p^2}, \quad \rho_X(z) = \frac{p}{1-(1-p)z}.$$

The number of Bernoulli trials $Y$ needed to get one success is also called a geometric distribution. Thus $Y = X + 1$, $S = \{1, 2, \ldots, \}$ and

$$f_Y(y|p) = p(1-p)^{y-1}.$$

# Negative Binomial Random Variables

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

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

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

- For the outcome $\{X = x\}$, the $r$-th success must occur on the $r+x$-th trial. So,
- we must have $r-1$ successes and $x$ failures in the first $r+x-1$ Bernoulli trials. This happens with probability

$$\binom{r+x-1}{x} p^{r-1} (1-p)^x.$$

- followed by success on the last trial. This happens with probability $p$.

# Negative Binomial Random Variables

The generating function

$$\rho_X(z) = \sum_{x=0}^{\infty} \binom{x+r-1}{x} p^r (1-p)^x z^x = p^r \sum_{x=0}^{\infty} \frac{(x+r-1)_x}{x!} \zeta^x = p^r (1-\zeta)^{-r}$$

where $\zeta = (1-p)z$, i.e., $\rho_X(z) = p^r(1-(1-p)z)^{-r}$.

Exercise. Check that the Taylor's series expansion of $g(\zeta) = (1-\zeta)^{-r}$ is the infinite sum given above. This gives the power series expansion of a negative power of the binomial. For this reason, $X$ is called a negative binomial random variable.

This shows that the negative binomial distribution is a mass function for any $r > 0$.

# Negative Binomial Random Variables

Exercise. Use the generating function to show that

$$EX = \frac{r(1-p)}{p} \quad \text{and} \quad \mathrm{Var}(X) = \frac{r(1-p)}{p^2}$$

$$\rho_X(z) = p^r(1-(1-p)z)^{-r}, \quad \rho_X'(z) = -rp^r(1-(1-p)z)^{-r-1}(-(1-p)),$$

$$EX = \rho_X'(1) = rp^r p^{-r-1}(1-p) = \frac{r(1-p)}{p}.$$

For integer values of $r$, a negative binomial random variable can be represented as the sum

$$X = Y_1 + \cdots + Y_r$$

of independent $Geom(p)$ random variables. This gives a second strategy to compute $EX$ and, as we shall soon learn, to compute $\mathrm{Var}(X)$

# 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.

- lottery tickets
  - $n$ is the number of lottery tickets sold and $p$ is the probability of winning,
- 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 $\lambda$ 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 $\lambda/n$ 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

$$
\begin{aligned}
P\{X=0\} &= \binom{n}{0}p^0(1-p)^n &&= \left(1-\frac{\lambda}{n}\right)^n \approx e^{-\lambda} \\
P\{X=1\} &= \binom{n}{1}p^1(1-p)^{n-1} &&= n\frac{\lambda}{n}\left(1-\frac{\lambda}{n}\right)^{n-1} \approx \lambda e^{-\lambda} \\
P\{X=2\} &= \binom{n}{2}p^2(1-p)^{n-2} &&= \frac{n(n-1)}{2}\left(\frac{\lambda}{n}\right)^2\left(1-\frac{\lambda}{n}\right)^{n-2} = \frac{n(n-1)}{n^2}\frac{\lambda^2}{2}\left(1-\frac{\lambda}{n}\right)^{n-2} \approx \frac{\lambda^2}{2}e^{-\lambda} \\
\vdots & \qquad\vdots && \qquad\qquad\qquad\vdots \\
P\{X=x\} &= \binom{n}{x}p^x(1-p)^{n-x} &&= \frac{(n)_x}{x!}\left(\frac{\lambda}{n}\right)^x\left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{(n)_x}{n^x}\frac{\lambda^x}{x!}\left(1-\frac{\lambda}{n}\right)^{n-x} \approx \frac{\lambda^x}{x!}e^{-\lambda}.
\end{aligned}
$$

Exercise. Explain why

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

# Poisson Random Variables

To compare the binomial to the Poisson, we look at an example with $\lambda = 2$.

| $x$ | $Bin(20, 0.1)$ | diff | $Bin(200, 0.01)$ | diff | $Pois(2)$ |
|---|---|---|---|---|---|
| 0 | 0.1216 | 0.0138 | 0.1340 | 0.0014 | 0.1353 |
| 1 | 0.2702 | 0.0005 | 0.2707 | 0.0000 | 0.2707 |
| 2 | 0.2852 | -0.0145 | 0.2720 | -0.0014 | 0.2707 |
| 3 | 0.1901 | -0.0097 | 0.1814 | -0.0009 | 0.1804 |
| 4 | 0.0898 | 0.0004 | 0.0902 | 0.0000 | 0.0902 |
| 5 | 0.0319 | 0.0042 | 0.0357 | 0.0004 | 0.0361 |
| 6 | 0.0089 | 0.0032 | 0.0117 | 0.0003 | 0.0120 |
| 7 | 0.0020 | 0.0015 | 0.0033 | 0.0002 | 0.0034 |
| | $\text{dist}_{TV}$ | $=0.0242$ | $\text{dist}_{TV}$ | $=0.0022$ | |

The total variation distance, $\text{dist}_{TV}$, is the maximum value among all events $A$ of the difference between the probabilities on that set $A$.

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

# Hypergeometric Random Variables

Let $Y_i$ be a Bernoulli random variable indicating whether or not the color of the $i$-th is white. Thus, its mean

$$EY_i = \frac{m}{m+n}.$$

The random variable $X = Y_1 + Y_2 + \cdots + Y_k$ and its mean

$$EX = EY_1 + EY_2 + \cdots + EY_k = k\frac{m}{m+n}.$$

We will later be able to compute the variance

$$\mathrm{Var}(X) = k\frac{m}{m+n} \cdot \frac{n}{m+n} \cdot \frac{m+n-k}{m+n-1}.$$

If we write $N = m + n$ and $p = m/(m+n)$, then

$$\mathrm{Var}(X) = kp(1-p)\frac{N-k}{N-1}.$$

# Summary

| random variable | R | parameters | mean | variance | generating function |
| --- | --- | --- | --- | --- | --- |
| Bernoulli | * | $p$ | $p$ | $p(1-p)$ | $(1-p)+pz$ |
| binomial | `binom` | $n, p$ | $np$ | $np(1-p)$ | $((1-p)+pz)^n$ |
| geometric | `geom` | $p$ | $\frac{1-p}{p}$ | $\frac{1-p}{p^2}$ | $\frac{p}{1-(1-p)z}$ |
| hypergeometric | `hyper` | $m, n, k$ | $\frac{km}{m+n}$ | $k\frac{m}{m+n} \cdot \frac{n}{m+n} \cdot \frac{m+n-k}{m+n-1}$ | |
| negative binomial | `nbinom` | $r, p$ | $r\frac{1-p}{p}$ | $r\frac{1-p}{p^2}$ | $\left(\frac{p}{1-(1-p)z}\right)^r$ |
| Poisson | `pois` | $\lambda$ | $\lambda$ | $\lambda$ | $\exp(-\lambda(1-z))$ |
| uniform | `sample` | $a, b$ | $\frac{b-a+1}{2}$ | $\frac{(b-a+1)^2-1}{12}$ | $\frac{z^a}{b-a+1}\frac{1-z^{b-a+1}}{1-z}$ |