Bernoulli distribution

The simplest non-uniform distribution

p – probability of success (1)

1-p – probability of failure (0)

$$f(x) = P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

Jacob Bernoulli
(1654-1705)
Swiss mathematician (Basel)

- Law of large numbers
- Mathematical constant e=2.718…

Bernoulli distribution

$$f(x) = P(X = x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \end{cases}$$

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = 0(1 - p) + 1(p) = p$$

$$\text{Var}(X) = E(X^2) - (EX)^2 = [0^2(1 - p) + 1^2(p)] - p^2 = p - p^2 = p(1 - p)$$

Refresher: Binomial Coefficients

$$\binom{n}{k} = C_k^n = \frac{n!}{k!(n-k)!}\ ,\ \text{called } n \text{ choose } k$$

$$\binom{10}{3} = C_3^{10} = \frac{10!}{3!7!} = \frac{10\cdot 9\cdot 8\cdot 7!}{3\cdot 2\cdot 1\cdot 7!} = 120$$

Number of ways to choose k objects out of n
without replacement and where the order does not matter.
Called binomial coefficients because of the binomial formula

$$(p+q)^n = (p+q)\times(p+q)\ldots\times(p+q) = \sum_{x=0}^{n} C_x^n p^x q^{n-x}$$

Binomial Distribution

- Binomially-distributed random variable X equals sum (number of successes) of n independent Bernoulli trials
- The probability mass function is:

$q = 1 - p$

$$f(x) = C_x^n p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \ldots n \qquad (3\text{-}7)$$

- Based on the binomial expansion:

$$1 = (p+q)^n = \sum_{x=0}^{n} C_x^n p^x q^{n-x}$$

Binomial Mean

X is a binomial random variable with parameters p and n

Mean:

$\mu = E(X) = np$

$$\mu = \sum x C_x^n p^x q^{n-x} = p \frac{\partial}{\partial p} \sum C_x^n p^x q^{n-x} =$$

$$= p \frac{\partial}{\partial p} (p+q)^n = np$$

$$E(X(X-1)) =$$

$$= \sum x(x-1) C^n_x p^x q^{n-x}$$

$$= p^2 \frac{\partial^2}{\partial p^2} \sum C^n_x p^x q^{n-x} =$$

$$= p^2 \frac{\partial^2}{\partial p^2} (p+q)^n \Big|_{q=1-p} = n(n-1)p^2$$

$$E(X^2) = E(X(X-1)) + E(X) =$$

$$= n^2 p^2 - np^2 + np = n^2 p^2 + np(1-p)$$

$$V(X) = E(X^2) - E(X)^2 = n^2 p^2 + np(1-p) - (np)^2$$

$$= \boxed{np(1-p)}$$

Binomial mean, variance and standard deviation

Let X be a binomial random variable with parameters p and n

- Mean:

$\mu = np$

- Variance:

$\sigma^2 = V(X) = np(1-p)$

- Standard deviation:

$\sigma = \sqrt{np(1-p)}$

- Standard deviation to mean ratio

$$\sigma/\mu = \sqrt{np(1-p)}/np = \frac{\sqrt{(1-p)/p}}{\sqrt{n}}$$

Matlab exercise: Binomial distribution

- Generate a sample of size 100,000 for binomially-distributed random variable X with n=100, p=0.2
- Tip: generate n Bernoulli random variables and use sum to add them up
- Plot the approximation to the Probability Mass Function based on this sample
- Calculate the mean and variance of this sample and compare it to theoretical calculations: $E[X]=n*p$ and $V[X]=n*p*(1-p)$

Matlab template: Binomial distribution

- **n=100; p=0.2;**
- **Stats=100000;**
- **r1=rand(Stats,n) ?? < or > ?? p;**
- **r2=sum(r1, ?? 1 rows or 2 columns ??);**
- **mean(r2)**
- **var(r2)**
- **[a,b]=hist(r2, 0:n);**
- **p_b=??./sum(??);**
- **figure; stem(??,p_b);**
- **figure; semilogy(??,p_b,'ko-')**

Matlab exercise: Binomial distribution

- **n=100; p=0.2;**
- **Stats=100000;**
- **r1=rand(Stats,n)<p;**
- **r2=sum(r1,2);**
- **mean(r2)**
- **var(r2)**
- **[a,b]=hist(r2, 0:n);**
- **p_b=a./sum(a);**
- **figure; stem(b,p_b);**
- **figure; semilogy(b,p_b,'ko-')**

Poisson Distribution

- Limit of the binomial distribution when
  - n , the number of attempts, is very large
  - p , *the probability of success* is very small
  - $E(X)=np=\lambda$ is O(1)

The annual numbers of deaths from horse kicks in 14 Prussian army corps between 1875 and 1894

Number of deaths	Observed frequency	Expected frequency
0	144	139
1	91	97
2	32	34
3	11	8
4	2	1
5 and over	0	0
Total	280	280

From von Bortkiewicz 1898

Siméon Denis Poisson (1781–1840)
French mathematician and physicist

Let $\lambda = np = E(x)$, so $p = \dfrac{\lambda}{n}$

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

$$= \frac{n(n-1)\ldots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} \sim \frac{n^x}{x!}\left(\frac{\lambda}{n}\right)^x = \frac{\lambda^x}{x!};$$

$$\sum_x \frac{\lambda^x}{x!} = e^{\lambda}.$$

Normalization requires $\sum_x P(X = x) = 1$.

Thus $P(X = x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$

Poisson Mean & Variance

If X is a Poisson random variable, then:

- Mean: $\mu = E(X) = \lambda$ $\simeq n \cdot p$
- Variance: $\sigma^2 = V(X) = \lambda$ $\simeq n \cdot p \cdot (1-p) \approx n \cdot p$
- Standard deviation: $\sigma = \lambda^{1/2}$

Note: Variance = Mean

Note: Standard deviation/Mean = $\lambda^{-1/2}$
decreases with λ