

Matlab exercise: Uniform distribution

- Generate a **sample of size 100,000** for uniform random variable X taking values $1,2,3,\dots,10$
- Plot the approximation to the **probability mass function** based on this sample
- Calculate mean and variance of this sample and compare it to **infinite sample predictions**:
 $E[X]=(a+b)/2$ and $V[X]=((a-b+1)^2-1)/12$

Matlab template: Uniform distribution

- `b=10; a=1; % b= upper bound; a= lower bound (inclusive)'`
- `Stats=100000; % sample size to generate`
- `r1=rand(Stats,1);`
- `r2=floor(??*r1)+??;`
- `mean(r2)`
- `var(r2)`
- `std(r2)`
- `[hy,hx]=hist(r2, 1:10); % hist generates histogram in bins 1,2,3...,10`
- `% hy - number of counts in each bin; hx - coordinates of bins`
- `p_f=hy./??; % normalize counts to add up to 1`
- `figure; plot(??,p_f, 'ko-'); ylim([0, max(p_f)+0.01]); % plot the PMF`

Matlab exercise: Uniform distribution

- `b=10; a=1; % b= upper bound; a= lower bound (inclusive)'`
- `Stats=100000; % sample size to generate`
- `r1=rand(Stats,1);`
- `r2=floor(b*r1)+a;`
- `mean(r2)`
- `var(r2)`
- `std(r2)`
- `[hy,hx]=hist(r2, 1:10); % hist generates histogram in bins 1,2,3...,10`
- `% hy - number of counts in each bin; hx - coordinates of bins`
- `p_f=hy./sum(hy); % normalize counts to add up to 1`
- `figure; plot(hx,p_f, 'ko-'); ylim([0, max(p_f)+0.01]); % plot the PMF`

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

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

Binomial Distribution

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

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

$q = 1-p$

- 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:

$$\mu = E(X) = np$$

$$\begin{aligned}\mu &= \sum x C_x^n p^x q^{n-x} = p \frac{\partial}{\partial p} \sum C_x^n p^x q^{n-x} = \\ &= p \frac{\partial}{\partial p} (p + q)^n = np\end{aligned}$$

$$\begin{aligned}
E(X(X-1)) &= \\
&= \sum x(x-1) C_x^n p^x q^{n-x} \\
&= p^2 \frac{\partial^2}{\partial p^2} \sum C_x^n p^x q^{n-x} = \\
&= p^2 \frac{\partial^2}{\partial p^2} (p+q)^n \Big|_{q=1-p} = n(n-1)p^2
\end{aligned}$$

$$\begin{aligned}
E(X^2) &= E(X(X-1)) + E(X) = \\
&= n^2 p^2 - n p^2 + n p = n^2 p^2 + n p (1-p)
\end{aligned}$$

$$\begin{aligned}
V(X) &= E(X^2) - E(X)^2 = n^2 p^2 + n p (1-p) - (np)^2 \\
&= \boxed{np(1-p)}
\end{aligned}$$

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-')`