Discrete Probability Functions

Daniel B. Rowe, Ph.D.

Professor
Department of Mathematical and Statistical Sciences

Outline

- Discrete RVs, PMFs, CDFs
- Discrete Expectations
- Discrete Moments
- Discrete Distributions

Discrete Random Variables

A random variable is a function from a sample space S into the real numbers X.

Examples:

Experiment	Random Variable
Toss two die	X=sum of numbers
Toss a coin 25 times	X= # heads in 25 tosses
Apply different amounts of fertilizer to plants	X=yield/acre

Discrete Probability Mass Functions

Example: Toss a coin twice.

$S=\{HH,HT,TH,TT\}$

X=number of heads.

Sample space outcomes

Real Numbers

$X(s)$ is a mapping from S to X.

$s_1=HH,\ s_2=HT,\ s_3=TH,\ s_4=TT$

s	HH	HT	TH	TT
X(s)	2	1	1	0

$X_1=2,\ X_2=1,\ X_3=1,\ X_4=0$

$P(X)$ is a mapping from X to [0,1].

x	0	1	2
$P(X=x)$	1/4	1/2	1/4

Real Numbers

Unit interval

Discrete Cumulative Distribution Functions

The *cdf* of a random variable X, $F_X(x)$, is

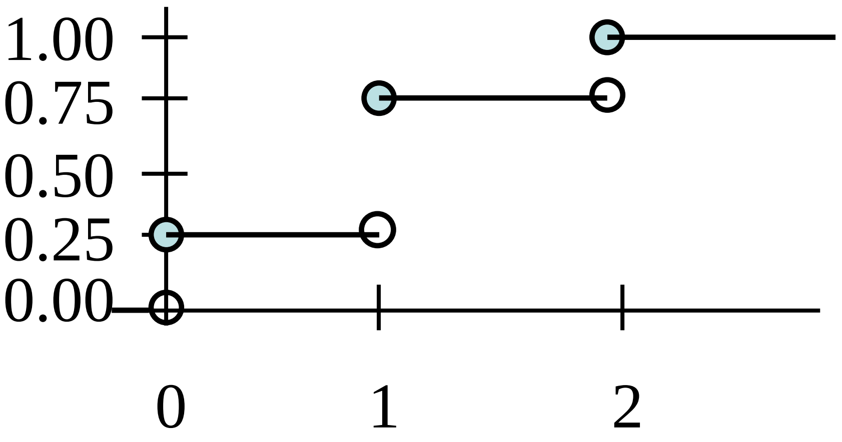
$$F_X(x) = P_X(X \le x), \text{ for all } x.$$

$$F_X(x) = \sum_{X \le x} P_X(X = x)$$

Example: Tossing a coin twice

x	0	1	2
P(X=x)	1/4	1/2	1/4

$$F_X(x) = \begin{cases} 0/4 & -\infty < x < 0 \\ 1/4 & 0 \le x < 1 \\ 3/4 & 1 \le x < 2 \\ 4/4 & 2 \le x < \infty \end{cases}$$



Discrete RVs, PMFs, and CDFs

Assume that the discrete random variable (RV) takes on values

$$x_1, x_2, x_3 \ldots$$

if the probability of each of value x_j is

$$P(X = x_j \mid \theta)$$

then, the probability mass function (PMF) is given by

$$f(x_j \mid \theta) = P(X = x_j \mid \theta) \text{ for } j = 1, 2, 3 \ldots$$

where θ are any parameters that the PMF depends on.

Discrete RVs, PMFs, and CDFs

Further, the cumulative distribution function (CDF) is given by

$$F(x|\theta) = \sum_{x_j \le x} f(x_j|\theta)$$

i.e. $F(x|\theta) = f(x_j|\theta) + \ldots + f(x_k|\theta)$

where x_k (largest value) ≤x.

Additionally, any PMF must satisfy

1) $0 \le f(x_j|\theta) \le 1$

2) $\sum_{x_j} f(x_j|\theta) = 1$.

Discrete Expectation

Given an arbitrary discrete probability mass $f(x_j|\theta)$, we want to compute quantitative population summaries of it such as

population mean, $\mu = \sum_{j=1}^{\infty} x_j f(x_j \mid \theta)$

population variance, $\sigma^2 = \sum_{j=1}^{\infty} (x_j - \mu)^2 f(x_j \mid \theta)$

population standard deviation, $\sigma = \sqrt{\sigma^2}$

Discrete Expectation

These population numerical summaries are found by expectation

$$E[g(X)\,|\,\theta]=\sum_{j=1}^{\infty} g(x_j)\,f(x_j\,|\,\theta) \quad .$$

The mean is

$$\mu = E(X\,|\,\theta)$$

and the variance is

$$\sigma^2 = E[(X-\mu)^2\,|\,\theta] \quad .$$

Discrete Expectation

Linearity Property:

$$E[\sum_{j=1}^{n} g_j(X) \mid \theta] = \sum_{j=1}^{n} E[g_j(X) \mid \theta]$$

i.e. $E[g_1(X) + g_2(X) \mid \theta] = E[g_1(X) \mid \theta] + E[g_2(X) \mid \theta]$

$$E[g_1(X) + g_2(X) \mid \theta] = \sum_{j=1}^{\infty} [g_1(x_i) + g_2(x_i)] f(x_i \mid \theta)$$

$$= \sum_{j=1}^{\infty} [g_1(x_j) f(x_j \mid \theta) + g_2(x_j) f(x_j \mid \theta)]$$

$$= \sum_{j=1}^{\infty} g_1(x_j) f(x_j \mid \theta) + \sum_{j=1}^{\infty} g_2(x_j) f(x_j \mid \theta)$$

Discrete Expectation

Example:

x	0	1	2
P(X=x)	1/4	1/2	1/4

$$E[g(X)\,|\,\theta] = \sum_{j=1}^{\infty} g(x_j)\, f(x_j\,|\,\theta)$$

$$g(X) = X$$

$$\mu = (0)(1/4) + (1)(2/4) + (2)(1/4)$$

$$\mu = 1$$

Discrete Expectation

Example:

x	0	1	2
P(X=x)	1/4	1/2	1/4

$$E[g(X)|\theta] = \sum_{j=1}^{\infty} g(x_j) f(x_j|\theta)$$

$$g(X) = (X-\mu)^2$$

$$\sigma^2 = (0-1)^2(1/4) + (1-1)^2(2/4) + (2-1)^2(1/4)$$

$$\sigma^2 = 1/2$$

Discrete Moments

For each integer n, the nth moment of X, μ_1' , is

$$\mu_n' = E(X^n \mid \theta) \quad .$$

The nth central moment of X, μ_n, is

$$\mu_n = E[(X - \mu)^n \mid \theta]$$

where $\mu_1' = E(X \mid \theta) \quad .$

Discrete Distributions

Uniform:

A random variable x has a discrete uniform distribution, x ~uniform(N) if

$$P(X = x \mid N) = \frac{1}{N}$$

where, N=1,2,… , x=1,2,…N.

i.e. flipping a coin, rolling a die.

Discrete Distributions

Uniform:

It can be shown that

$$\mu = \sum_{j=1}^{\infty} x_j P(X = x_j) = \sum_{j=1}^{N} j \frac{1}{N} = \frac{N+1}{2}$$

Note:

$$1 + 2 + 3 + ... + N = \frac{N(N+1)}{2}$$

Discrete Distributions

Uniform:

that

$$\sigma^2 = \sum_{j=1}^{\infty} \left(x_j - \mu\right)^2 P(X = x_j)$$

$$= \sum_{j=1}^{N} \left(j - \frac{N+1}{2} \right)^2 \frac{1}{N}$$

$$= \frac{N^2 - 1}{12}$$

Note:

$$1 + 2^2 + 3^2 + \ldots + N^2 = \frac{N(N+1)(2N+1)}{6}$$

Discrete Distributions

Uniform:

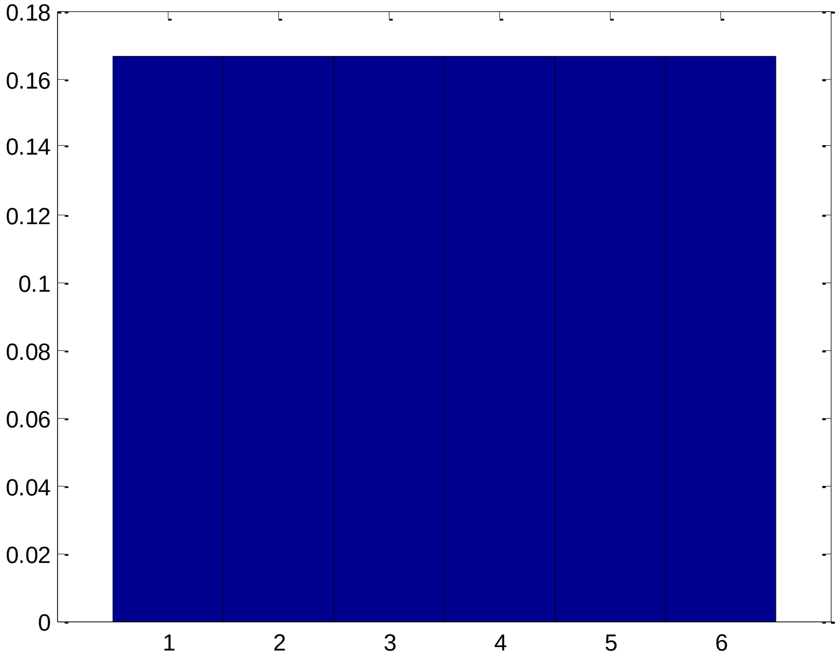
$$P(X = x \mid N) = \frac{1}{N}$$ where, N=1,2,… , x=1,2,…N.

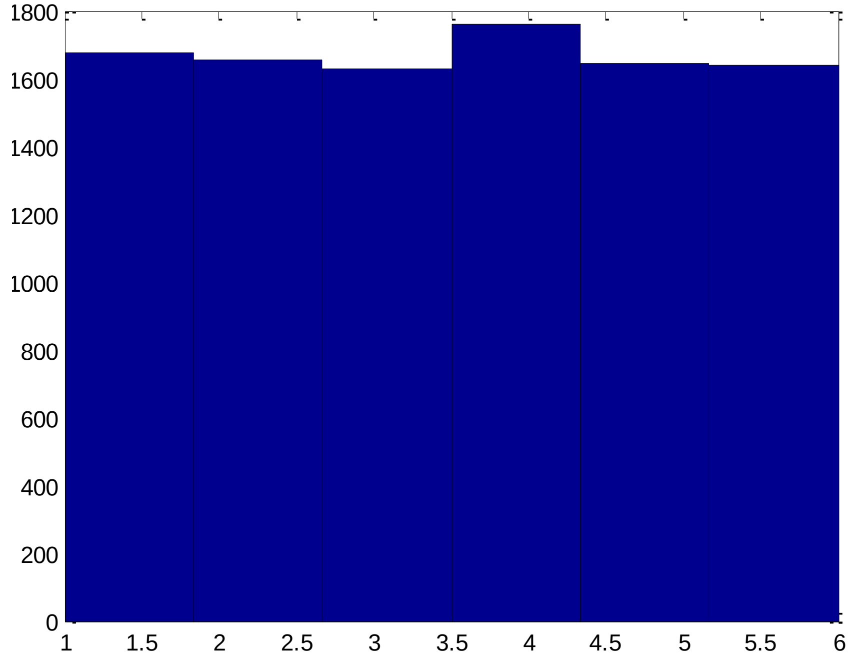
Uniform discrete random variates can be generated by either partitioning the unit interval into N bins, generating uniform [0,1] numbers and for each random number convert to j if it is in the j^{th} bin or

```
N=6;
x=unidrnd(N)
5
```

Discrete Distributions

Uniform: $P(X = x) = \frac{1}{6}$



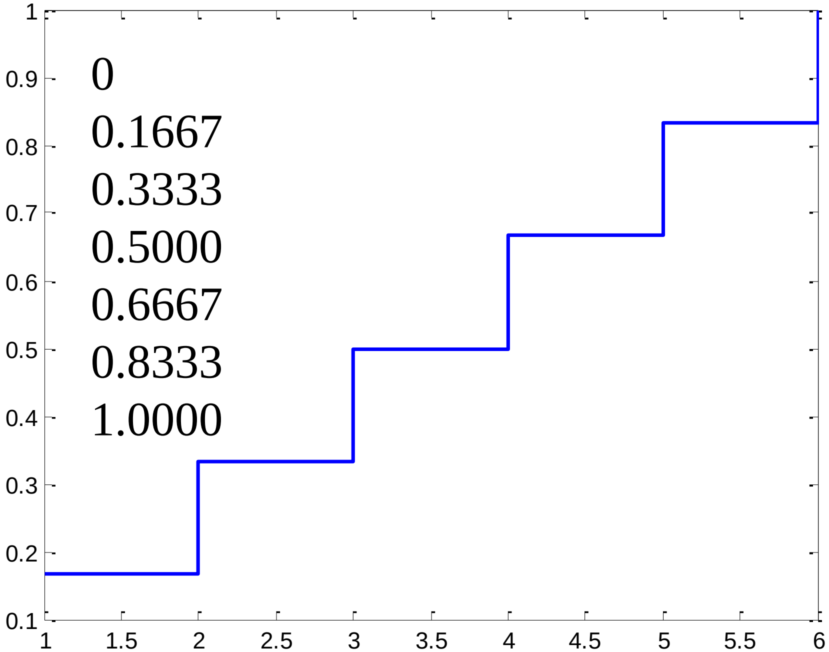



```
N=6;
xpop=(1:N);
ypop = unidpdf(x,N);
bar(xpop,ypop,1)
```

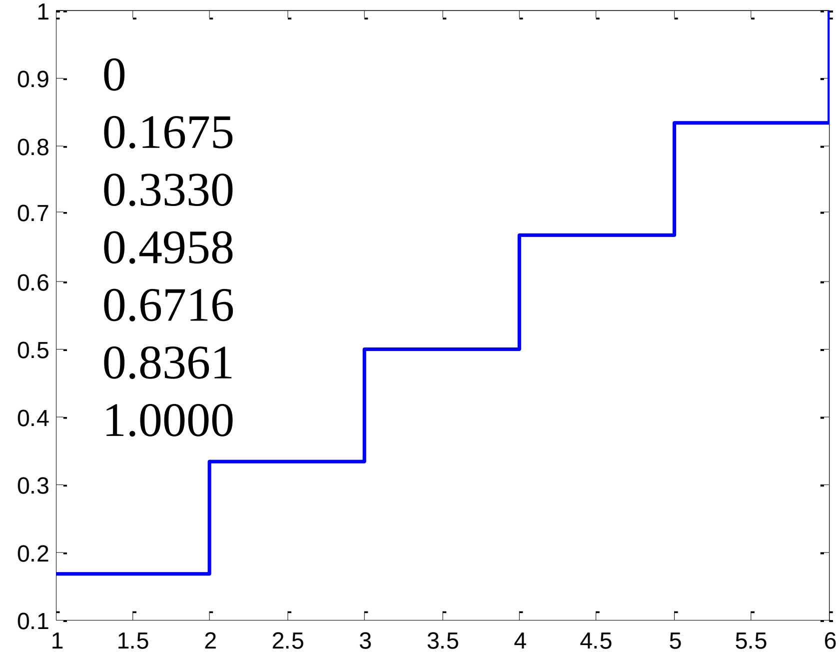
```
N=6;, num=10000;
xsamp=unidrnd(N,num,1);
hist(xsamp,N)
```

Discrete Distributions

Uniform: $P(X=x)=\frac{1}{6}$




```
ducdf =unidcdf(x,6)
stairs(x, ducdf,'LineWidth',2)
```




```
[F,x]=ecdf(xsamp);
stairs(x,F,'LineWidth',2)
```

Discrete Distributions

Uniform: $P(X=x)=\frac{1}{6}$

unidstat(N)

mean(xsamp)
var(xsamp)

	True	Simulated
μ	3.5	3.4960
σ^2	2.9167	2.8987

Discrete Distributions

Bernoulli:

A random variable x has a discrete Bernoulli distribution, x~Bernoulli(p) if

$$P(X=x\,|\,p)=p^x(1-p)^{1-x}$$

where, $0\leq p\leq 1$, $x=0,1$.

x = the number of successes out of 1 trial.

p = the probability of success on the trial.

Discrete Distributions

Bernoulli:

It can be shown that

$$\begin{aligned}\mu &= \sum_{j=1}^{\infty} x_j P(X = x_j) \\ &= (0)(1-p)+(1)p \\ &= p\end{aligned}$$

Discrete Distributions

Bernoulli:

that

$$\sigma^2 = \sum_{j=1}^{\infty} \left(x_j - \mu \right)^2 P(X = x_j)$$

$$= \left(0 - p\right)^2 (1 - p) + \left(1 - p\right)^2 p$$

$$= p(1 - p)$$

Discrete Distributions

Bernoulli:

$P(X = x \mid p) = p^x (1-p)^{1-x}$ where, $0 \le p \le 1$, $x = 0,1$.

Bernoulli discrete random variates can be generated by either partitioning the unit interval into 2 bins, generating uniform [0,1] numbers and for each random number convert to j-1 if it is in the j^{th} bin or

```
p=1/2; n=1;
x=binornd(1,p)
1
```

Discrete Distributions

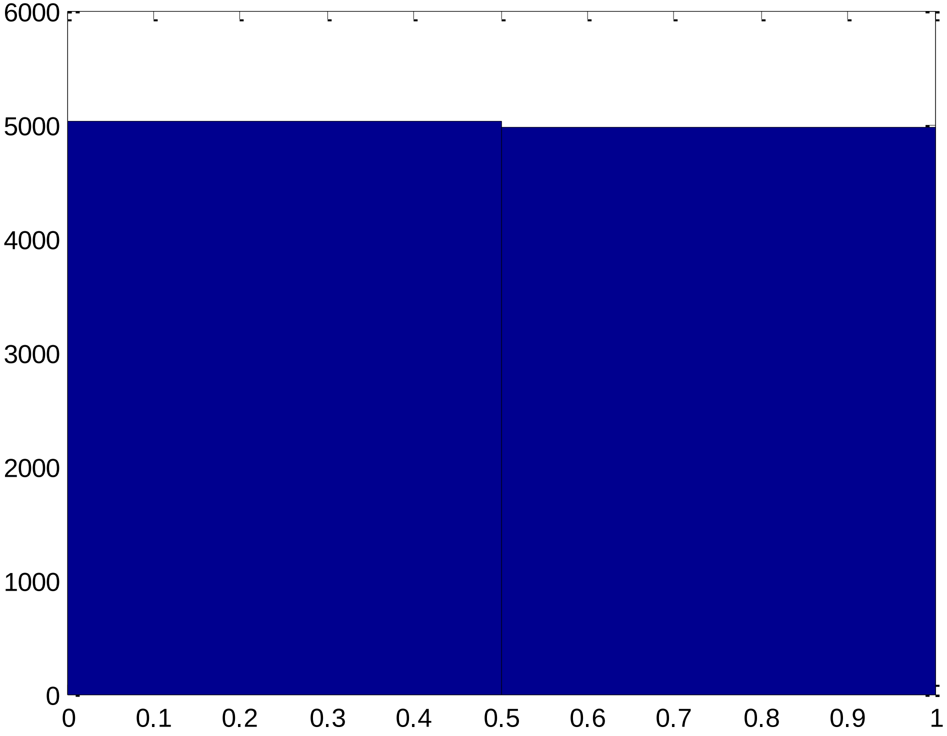
Bernoulli:

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

```
n=1; p=1/2; num=10^4;
x=binornd (n,p,num,1);
mean(x)
var(x)
hist(x,2)
```

	True	Simulated
μ	0.5	0.4971
σ^2	0.25	0.2500

Can also find and plot ECDF.



Discrete Distributions

Binomial:

For the Bernoulli distribution, $P(X=x\,|\,p)=p^{x}(1-p)^{1-x}$.

Let X_1~Bernoulli(p) and X_2~Bernoulli(p),

Define $X=X_1+X_2$. i.e. X=total number of successes

The distribution of X is binomial(2,p).

This can be repeated to get binomial(n,p).

Discrete Distributions

Binomial:

A random variable x has a discrete binomial distribution, x~binomial(n,p) if

$$P(X=x\,|\,n,p)=\underbrace{\frac{n!}{x!(n-x)!}}_{n(x\text{ successes})}\underbrace{p^x(1-p)^{n-x}}_{\substack{P(x\text{ successes}\\ \text{and } n-x\text{ failures})}}$$

where, $0\le p\le 1$, $n=1,2,3,\ldots$, $x=0,1\ldots,n$.

n = number of trials or times we repeat the experiment.

x = the number of successes out of n trials.

p = the probability of success on an individual trial.

Discrete Distributions

Binomial:

It can be shown that

$$\begin{aligned} \mu &= \sum_{j=1}^{\infty} x_j P(X = x_j) \\ &= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \\ &= np \end{aligned}$$

Discrete Distributions

Binomial:

that

$$\sigma^2 = \sum_{j=1}^{\infty} \left(x_j - \mu\right)^2 P(X = x_j)$$

$$= \sum_{x=0}^{n} (x - np)^2 \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

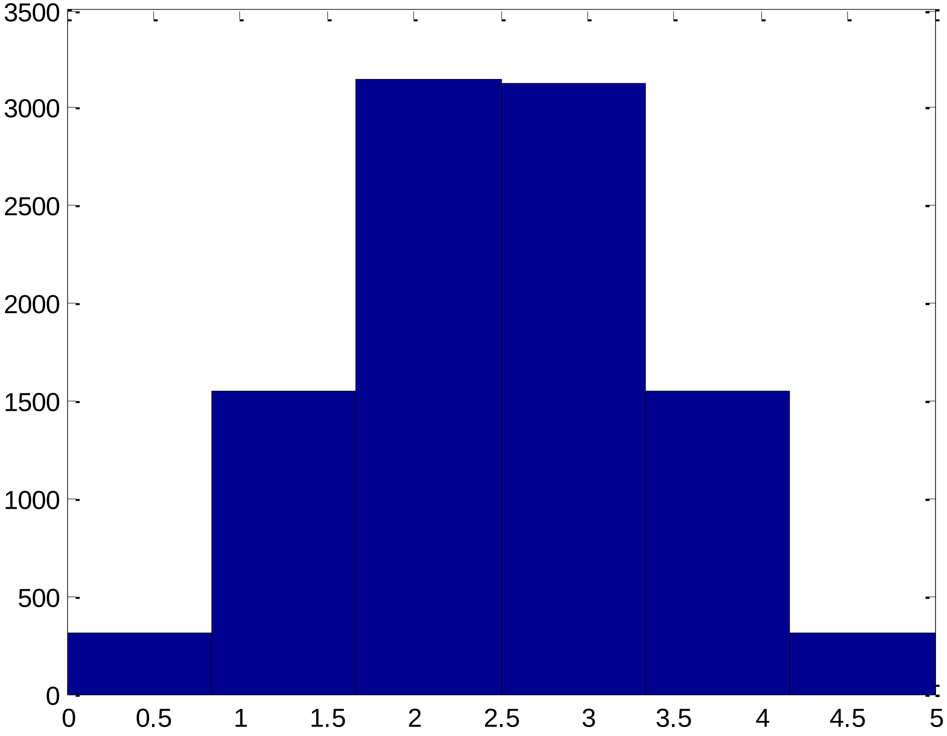
$$= np(1-p)$$

Discrete Distributions

Binomial: $$P(X=x\,|\,n,p)=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

```
n=5; p=1/2;num=10^4;
x=binornd (n,p,num,1);
mean(x)
var(x)
hist(x,6)
```

	True	Simulated
μ	2.5	2.4982
σ^2	1.25	1.2455



Can also find and plot ECDF.

Discrete Distributions

Poisson:

A random variable x has a discrete Poisson distribution, x~Poisson(λ) if

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

where, $\lambda > 0$, $x = 0,1\ldots$.

λ is called the intensity parameter

Discrete Distributions

Poisson:

It can be shown that

$$\begin{aligned} \mu &= \sum_{j=1}^{\infty} x_j P(X = x_j) \\ &= \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} \\ &= \lambda \end{aligned}$$

Discrete Distributions

Poisson:

that

$$\sigma^2 = \sum_{j=1}^{\infty} \left(x_j - \mu\right)^2 P(X = x_j)$$

$$= \sum_{x=0}^{\infty} (x - \lambda)^2 \frac{\lambda^x e^{-\lambda}}{x!}$$

$$= \lambda$$

Discrete Distributions

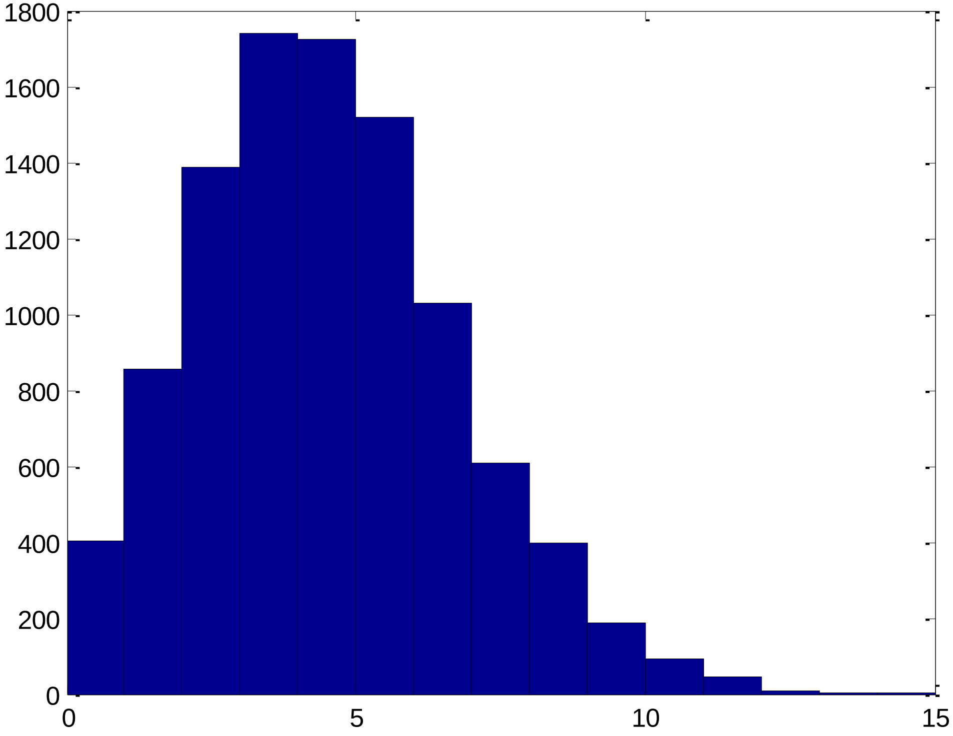
Poisson:

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

```
lam=5;num=10^4;
x=poissrnd(lam,num,1);
mean(x)
var(x)
hist(x,15)
```

	True	Simulated
μ	5	5.0181
σ^2	5	5.1069

Can also find and plot ECDF.



Discrete Distributions

Geometric:

A random variable x has a discrete geometric distribution, x~geometric(p) if

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

where, $0 \le p \le 1$, $x = 1, 2 \ldots$.

x = the trial at which first successes occurs.

p = the probability of success on the trial.

Discrete Distributions

Geometric:

It can be shown that

$$\begin{aligned} \mu &= \sum_{j=1}^{\infty} x_j P(X = x_j) \\ &= \sum_{x=1}^{\infty} xp(1-p)^{x-1} \\ &= \frac{1}{p} \end{aligned}$$

$$q = (1-p)$$

Discrete Distributions

Geometric:

that

$$
\begin{aligned}
\sigma^2 &= \sum_{j=1}^{\infty} \left(x_j - \mu\right)^2 P(X = x_j) \\
&= \sum_{x=1}^{\infty} (x - \frac{1}{p})^2 p(1-p)^{x-1} \\
&= \frac{q}{p^2}
\end{aligned}
$$

$$q = (1-p)$$

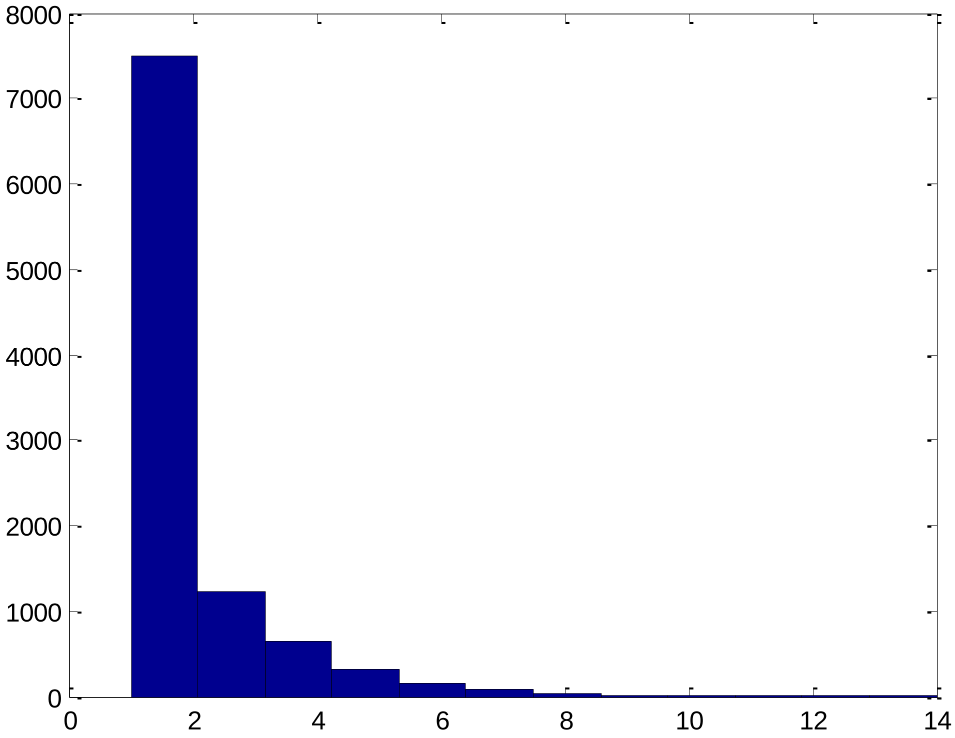
Discrete Distributions

Geometric:

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

```
p=1/2;num=10^4;
x= geornd(p,num,1)+1;
mean(x)
var(x)
hist(x,12)
```

	True	Simulated
μ	2	1.9994
σ^2	2	2.0358



Can also find and plot ECDF.

Homework 2:

1) Generate 10^6 observations from the discrete uniform, Bernoulli, binomial, Poisson, and geometric distributions. Use the parameter values I used.

   Make a histogram. Compute sample mean and variance. Compare to what true values are.
   Compute an empirical CDF and find 50^{th} percentile.
   Compute and plot an incremental sample mean and variance. When do you think you get convergence?