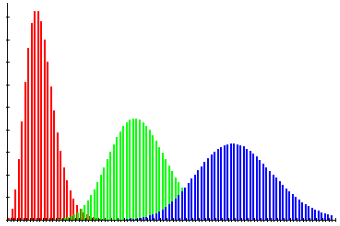
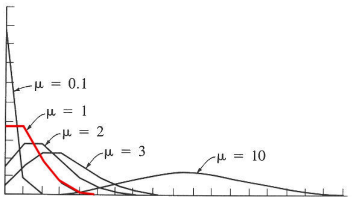
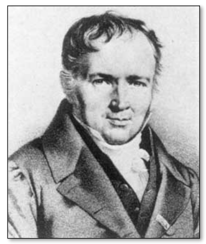


a zoo of (discrete) random variables





discrete uniform random variables

A discrete random variable X equally likely to take any (integer) value between integers a and b, inclusive, is *uniform.*

Notation: $X \sim \text{Unif}(a,b)$

Probability: $P(X=i) = \dfrac{1}{b-a+1}$

Mean, Variance: $E[X] = \dfrac{a+b}{2}, \text{Var}[X] = \dfrac{(b-a)(b-a+2)}{12}$

discrete uniform random variables

A discrete random variable X equally likely to take any (integer) value between integers a and b, inclusive, is *uniform.*

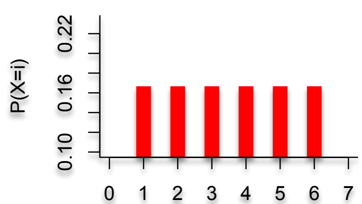
Notation: $X \sim \text{Unif}(a,b)$

Probability: $P(X=i) = \dfrac{1}{b-a+1}$

Mean, Variance: $E[X] = \dfrac{a+b}{2}, \text{Var}[X] = \dfrac{(b-a)(b-a+2)}{12}$

Example: value shown on one roll of a fair die is Unif(1,6):

$P(X=i) = 1/6$
$E[X] = 7/2$
$\text{Var}[X] = 35/12$



Bernoulli random variables

An experiment results in "Success" or "Failure"

X is an *indicator random variable* (1 = success, 0 = failure)

$P(X=1) = p$ and $P(X=0) = 1-p$

X is called a *Bernoulli* random variable: $X \sim \text{Ber}(p)$

$E[X] = E[X^2] = p$

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

Bernoulli random variables

An experiment results in "Success" or "Failure"

- X is an *indicator random variable* (1 = success, 0 = failure)
  - P(X=1) = p   and   P(X=0) = 1-p
- X is called a *Bernoulli* random variable: X ~ Ber(p)

Examples:

- coin flip
- random binary digit
- whether a disk drive crashed

binomial random variables

Consider n independent random variables Y_i ~ Ber(p)

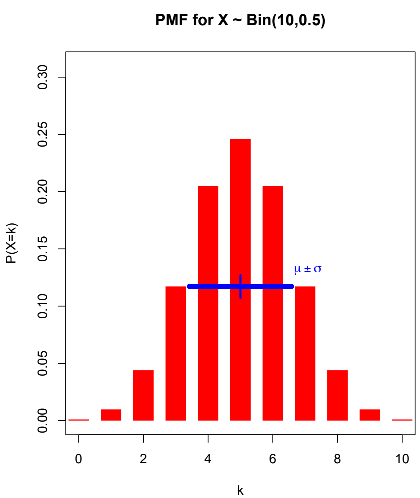
- X = $\Sigma_i Y_i$ is the number of successes in n trials
- X is a *Binomial* random variable: X ~ Bin(n,p)

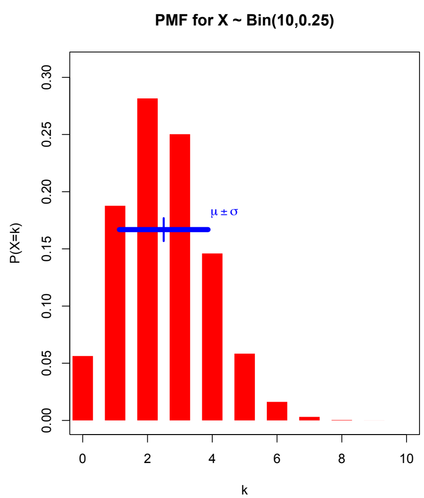
$$P(X = i) = \binom{n}{i} p^i (1-p)^{n-i} \quad i = 0, 1, \ldots, n$$

Examples

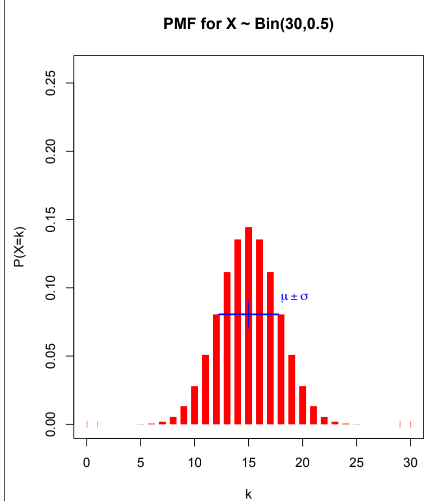
- # of heads in n coin flips
- # of 1's in a randomly generated length n bit string
- # of disk drive crashes in a 1000 computer cluster

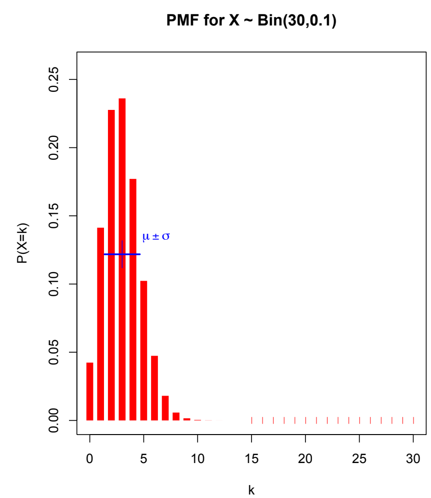
binomial pmfs





binomial pmfs





mean, variance of the binomial (II)

If $Y_1, Y_2, \ldots, Y_n \sim \text{Ber}(p)$ and independent,

then $X = \sum_{i=1}^{n} Y_i \sim \text{Bin}(n,p)$.

$\boxed{E[X] = np}$

$$E[X] = E\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} E\left[Y_i\right] = nE[Y_1] = np$$

$\boxed{\text{Var}[X] = np(1-p)}$

$$\text{Var}[X] = \text{Var}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \text{Var}\left[Y_i\right] = n\text{Var}[Y_1] = np(1-p)$$

geometric distribution

In a series $X_1, X_2, \ldots$ of Bernoulli trials with success probability p, let Y be the index of the first success, i.e.,

$X_1 = X_2 = \ldots = X_{Y-1} = 0$ & $X_Y = 1$

Then Y is a *geometric* random variable with parameter p.

Examples:

Number of coin flips until first head

Number of blind guesses on SAT until I get one right

Number of darts thrown until you hit a bullseye

Number of random probes into hash table until empty slot

Number of wild guesses at a password until you hit it

geometric distribution

In a series $X_1, X_2, \ldots$ of Bernoulli trials with success probability p, let Y be the index of the first success, i.e.,

$X_1 = X_2 = \ldots = X_{Y-1} = 0$ & $X_Y = 1$

Then Y is a *geometric* random variable with parameter p.

$P(Y=k) = (1-p)^{k-1}p$;   Mean $1/p$;   Variance $(1-p)/p^2$

geometric distribution

Flip a (biased) coin repeatedly until 1st head observed

How many flips? Let X be that number.

$P(X=1) = P(H) = p$

$P(X=2) = P(TH) = (1-p)p$

$P(X=3) = P(TTH) = (1-p)^2p$

...

$$\boxed{\sum_{i\geq 0} x^i = \frac{1}{1-x}, \text{ when } |x| < 1}$$

memorize me!

Check that it is a valid probability distribution:

1) $\forall i \geq 1, P(\{X = i\}) \geq 0$

2) $$P\left(\bigcup_{i\geq 1}\{X = i\}\right) = \sum_{i\geq 1}(1-p)^{i-1}p = p\sum_{i\geq 0}(1-p)^i = p\frac{1}{1-(1-p)} = 1$$

geometric random variable Geo(p)

Let X be the number of flips up to & including 1st head observed in repeated flips of a biased coin.

$$P(H) = p; \quad P(T) = 1 - p = q$$
$$p(i) = pq^{i-1} \quad \leftarrow \textit{PMF}$$
$$E[X] = \sum_{i \geq 1} ip(i) = \sum_{i \geq 1} ipq^{i-1} = p \sum_{i \geq 1} iq^{i-1} \quad (*)$$

A calculus trick:

$$\sum_{i \geq 1} iy^{i-1} = \sum_{i \geq 1} \frac{d}{dy} y^i = \sum_{i \geq 0} \frac{d}{dy} y^i = \frac{d}{dy} \sum_{i \geq 0} y^i = \frac{d}{dy} \frac{1}{1-y} = \frac{1}{(1-y)^2}$$

$dy^0/dy = 0$

So (*) becomes:

$$E[X] = p \sum_{i \geq 1} iq^{i-1} = \frac{p}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p}$$

E.g.:

p=1/2; on average head every 2nd flip

p=1/10; on average, head every 10th flip.

models & reality

Sending a bit string over the network

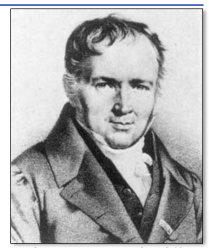
- n = 4 bits sent, each corrupted with probability 0.1
- X = # of corrupted bits, X ~ Bin(4, 0.1)
- In real networks, large bit strings (length n ≈ 10^4)
- Corruption probability is very small: p ≈ 10^{-6}
- X ~ Bin(10^4, 10^{-6}) is unwieldy to compute

Extreme n and p values arise in many cases

- # bit errors in file written to disk
- # of typos in a book
- # of elements in particular bucket of large hash table
- # of server crashes per day in giant data center
- # facebook login requests sent to a particular server

Poisson random variables

Suppose "events" happen, independently, at an *average* rate of λ per unit time. Let X be the *actual* number of events happening in a given time unit. Then X is a *Poisson* r.v. *with parameter* λ (denoted X ~ Poi(λ)) and has distribution (PMF):



Siméon Poisson, 1781-1840

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

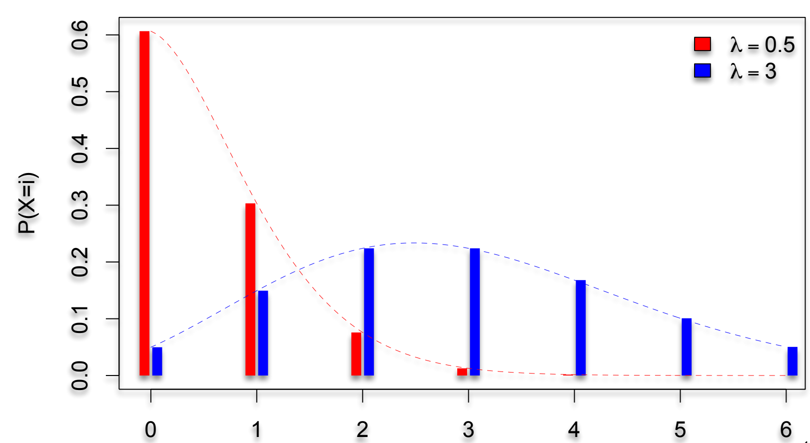
Examples:

- # of alpha particles emitted by a lump of radium in 1 sec.
- # of traffic accidents in Seattle in one year
- # of babies born in a day at UW Med center
- # of visitors to my web page today

See B&T Section 6.2 for more on theoretical basis for Poisson.

poisson random variables

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



poisson random variables

X is a Poisson r.v. with parameter λ if it has PMF:

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

Is it a valid distribution? Recall Taylor series:

$$e^\lambda = \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \cdots = \sum_{0 \le i} \frac{\lambda^i}{i!}$$

So

$$\sum_{0 \le i} P(X = i) = \sum_{0 \le i} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{0 \le i} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$$

poisson random variables

X is a Poisson r.v. with parameter λ if it has PMF:

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

Is it a valid distribution? Recall Taylor series:

$$e^\lambda = \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \cdots = \sum_{0 \le i} \frac{\lambda^i}{i!}$$

So

$$\sum_{0 \le i} P(X = i) = \sum_{0 \le i} e^{-\lambda} \frac{\lambda^i}{i!} = e^{-\lambda} \sum_{0 \le i} \frac{\lambda^i}{i!} = e^{-\lambda} e^{\lambda} = 1$$

expected value of poisson r.v.s

$$\begin{aligned}
E[X] &= \sum_{0 \le i} i \cdot e^{-\lambda} \frac{\lambda^i}{i!} \\
&= \sum_{1 \le i} i \cdot e^{-\lambda} \frac{\lambda^i}{i!} \\
&= \lambda e^{-\lambda} \sum_{1 \le i} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} \sum_{0 \le j} \frac{\lambda^j}{j!} \\
&= \lambda e^{-\lambda} e^{\lambda} \\
&= \lambda
\end{aligned}$$

i = 0 term is zero

j = i-1

As expected, given definition in terms of "average rate λ"

(Var[X] = λ, too; proof similar, see B&T example 6.20)

binomial —> Poisson in the limit

X ~ Binomial(n,p)

Poisson approximates binomial when n is large, p is small, and λ = np is "moderate"

binomial → poisson in the limit

X ~ Binomial(n,p)

$$
\begin{aligned}
P(X=i) &= \binom{n}{i} p^i (1-p)^{n-i} \\
&= \frac{n!}{i!(n-i)!}\left(\frac{\lambda}{n}\right)^i \left(1-\frac{\lambda}{n}\right)^{n-i}, \text{ where } \lambda = pn \\
&= \frac{n(n-1)\cdots(n-i+1)}{n^i}\,\frac{\lambda^i}{i!}\,\frac{(1-\lambda/n)^n}{(1-\lambda/n)^i} \\
&= \underbrace{\frac{n(n-1)\cdots(n-i+1)}{(n-\lambda)^i}}\,\frac{\lambda^i}{i!}\,\underbrace{(1-\lambda/n)^n} \\
&\approx \quad 1 \quad \cdot \frac{\lambda^i}{i!} \cdot \quad e^{-\lambda}
\end{aligned}
$$

I.e., Binomial ≈ Poisson for large n, small p, moderate i, λ.

Handy: Poisson has only 1 parameter–the expected # of successes

binomial random variable is poisson in the limit

Poisson approximates binomial when n is large, p is small, and λ = np is "moderate"

Different interpretations of "moderate," e.g.

- n > 20 and p < 0.05
- n > 100 and p < 0.1

Formally, Binomial is Poisson in the limit as n → ∞ (equivalently, p → 0) while holding np = λ

sending data on a network

Consider sending bit string over a network

- Send bit string of length n = 10^4
- Probability of (independent) bit corruption is p = 10^{-6}
- X ~ Poi(λ = $10^4 \cdot 10^{-6}$ = 0.01)
- What is probability that message arrives uncorrupted?

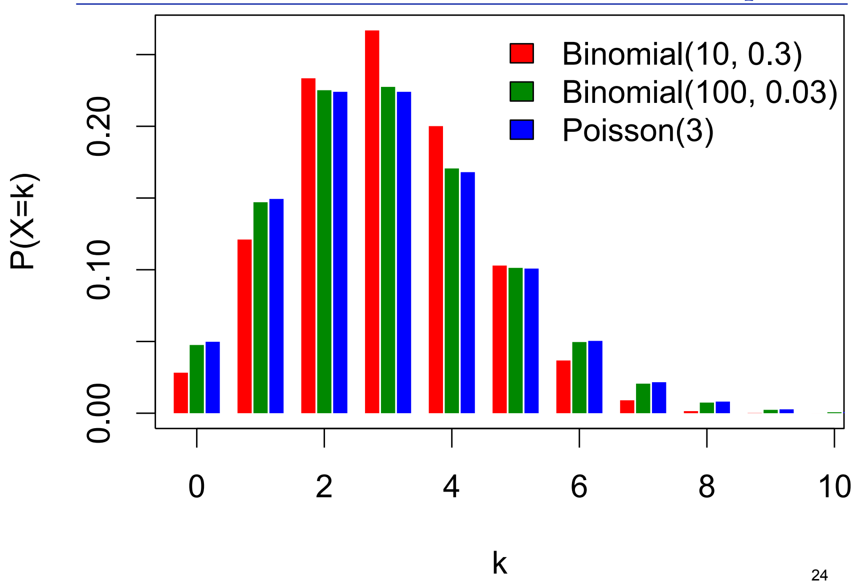
$$P(X=0) = e^{-\lambda}\frac{\lambda^0}{0!} = e^{-0.01}\frac{0.01^0}{0!} \approx 0.990049834$$

Using Y ~ Bin(10^4, 10^{-6}):

P(Y=0) ≈ 0.990049829

I.e., Poisson approximation (here) is accurate to ~5 parts per billion

binomial vs poisson



expectation and variance of a poisson

Recall: if Y ~ Bin(n,p), then:

E[Y] = pn

Var[Y] = np(1-p)

And if X ~ Poi(λ) where λ = np (n →∞, p → 0) then

E[X] = λ = np = E[Y]

Var[X] = λ ≈ λ(1-λ/n) = np(1-p) = Var[Y]

expectation and variance of a poisson

Recall: if Y ~ Bin(n,p), then:

E[Y] = pn

Var[Y] = np(1-p)

And if X ~ Poi(λ) where λ = np (n →∞, p → 0) then

E[X] = λ = np = E[Y]

Var[X] = λ ≈ λ(1-λ/n) = np(1-p) = Var[Y]

Expectation and variance of Poisson are the same (λ)

Expectation is the same as corresponding binomial

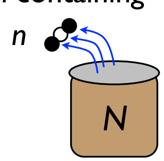
Variance almost the same as corresponding binomial

Note: when two different distributions share the same mean & variance, it suggests (but doesn't prove) that one may be a good approximation for the other.

balls in urns – the hypergeometric distribution

Draw *n* balls (without replacement) from an urn containing *N*, of which *m* are white, the rest black.

Let *X* = number of white balls drawn



$$P(X = i) = \frac{\binom{m}{i}\binom{N-m}{n-i}}{\binom{N}{n}}$$

E[X] = np, where p = m/N (the fraction of white balls)

proof: Let X_j be 0/1 indicator for j-th ball is white, X = Σ X_j

The X_j are *dependent*, but E[X] = E[Σ X_j] = Σ E[X_j] = np

Var[X] = np(1-p)(1-(n-1)/(N-1))

like binomial (almost)

some important (discrete) distributions

Name	PMF	E(X)	E(X²)	σ^2
Uniform(a,b)	$f(k) = \frac{1}{(b-a+1)}, k = a, a+1, \ldots, b$	$\frac{a+b}{2}$		$\frac{(b-a+1)^2-1}{12}$
Bernoulli(p)	$f(k) = \begin{cases} 1-p & \text{if } k=0 \\ p & \text{if } k=1 \end{cases}$	p	p	$p(1-p)$
Binomial(p,n)	$f(k) = \binom{n}{k}p^k(1-p)^{n-k}, k = 0,1,\ldots,n$	np		$np(1-p)$
Poisson(λ)	$f(k) = e^{-\lambda}\frac{\lambda^k}{k!}, k = 0,1,\ldots$	λ	$\lambda(\lambda+1)$	λ
Geometric(p)	$f(k) = p(1-p)^{k-1}, k = 1,2,\ldots$	$\frac{1}{p}$	$\frac{2-p}{p^2}$	$\frac{1-p}{p^2}$
Hypergeometric(n,N,m)	$f(k) = \frac{\binom{m}{k}\binom{N-m}{n-k}}{\binom{N}{n}}, k = 0,1,\ldots,N$	$\frac{nm}{N}$		$\frac{nm}{N}\left(\frac{(n-1)(m-1)}{N-1} + 1 - \frac{nm}{N}\right)$