

1.10.2 Normal distribution
1.10.3 Approximating binomial distribution by
normal
2.10 Central Limit Theorem

Prof. Tesler

Math 283
Fall 2019

Normal distribution

a.k.a. “Bell curve” and “Gaussian distribution”

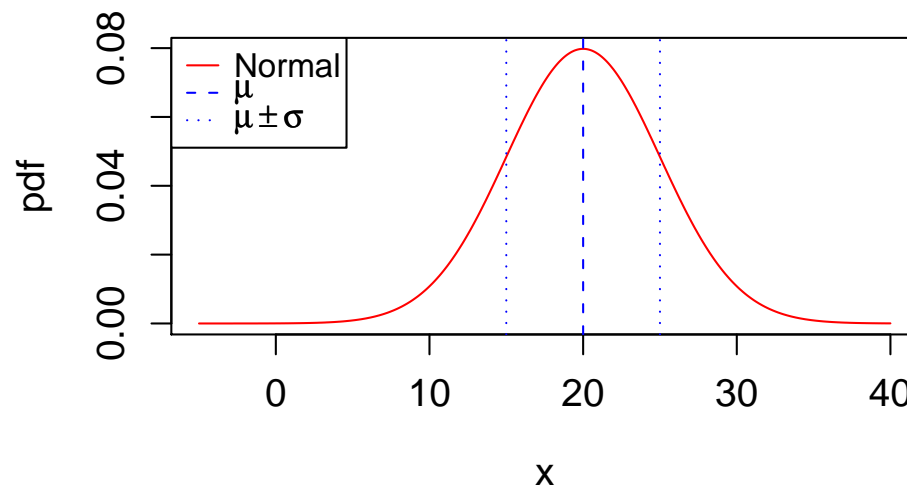
- The *normal distribution* is a continuous distribution. Parameters:

μ = mean (center)

σ = standard deviation (width)

- PDF: $f_X(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ for $-\infty < x < \infty$.

Normal distribution $N(20, 5)$: $\mu = 20$, $\sigma = 5$



- The normal distribution is symmetric about $x = \mu$, so median = mean = μ .

Applications of normal distribution

Applications

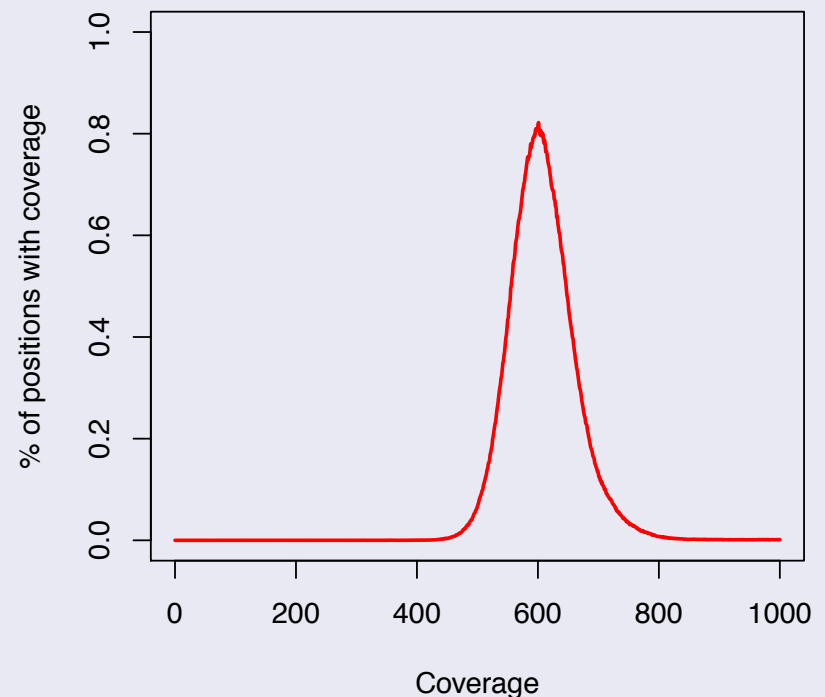
- Many natural quantities are modelled by it: e.g., a histogram of the heights or weights of everyone in a large population often follows a normal distribution.
- Many distributions such as binomial, Poisson,... are closely approximated by it when the parameters are large enough.
- Sums and averages of huge quantities of data are often modelled by it.

Coverage in DNA sequencing

Illumina GA_{II} sequencing of *E. coli* at 600× coverage.

Chitsaz et al. (2011), *Nature Biotechnology*

Empirical distribution of coverage



Cumulative distribution function

- The cumulative distribution function is the integral

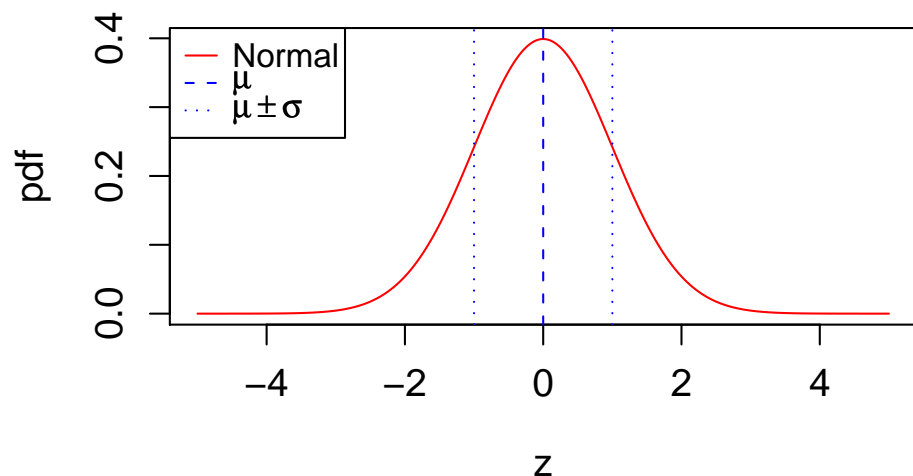
$$F_X(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(t - \mu)^2}{2\sigma^2}\right) dt$$

- The usual strategy to compute integrals is antiderivatives, like $\int x^2 dx = \frac{x^3}{3} + C$. But this doesn't have an antiderivative in terms of the usual functions (polynomials, exponentials, logs, trig, ...).
- The integral can be done via numerical integration or Taylor series.
- The integral for total probability equals 1; this can be shown using double integrals in polar coordinates:

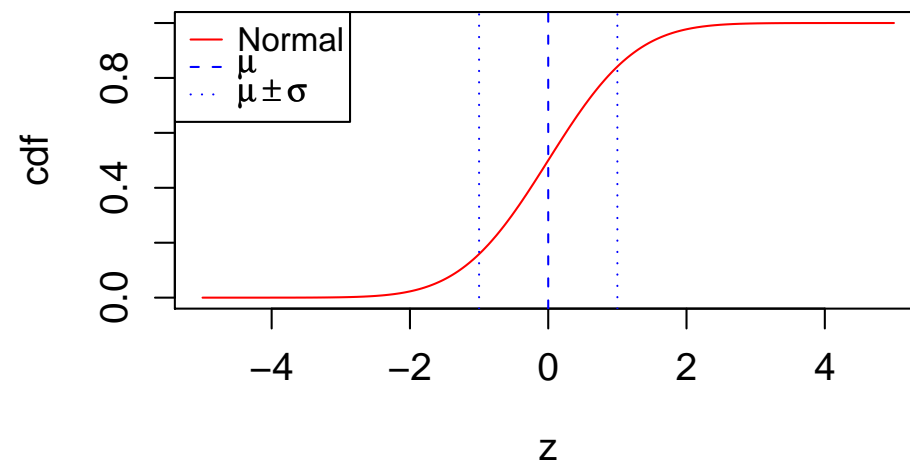
$$\int_{-\infty}^{\infty} \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx = 1$$

Standard normal distribution

Standard normal distribution $N(0, 1)$: $\mu = 0$, $\sigma = 1$



CDF of standard normal distribution



- The **standard normal distribution** is the normal distribution for $\mu = 0$, $\sigma = 1$. Use the variable name Z :

$$\text{PDF: } \phi(z) = f_Z(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}} \quad \text{for } -\infty < z < \infty$$

$$\text{CDF: } \Phi(z) = F_Z(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-t^2/2} dt$$

- The integral requires numerical methods. In the past, people used lookup tables. We'll use functions for it in Matlab and R.

Matlab and R commands

- For the standard normal:

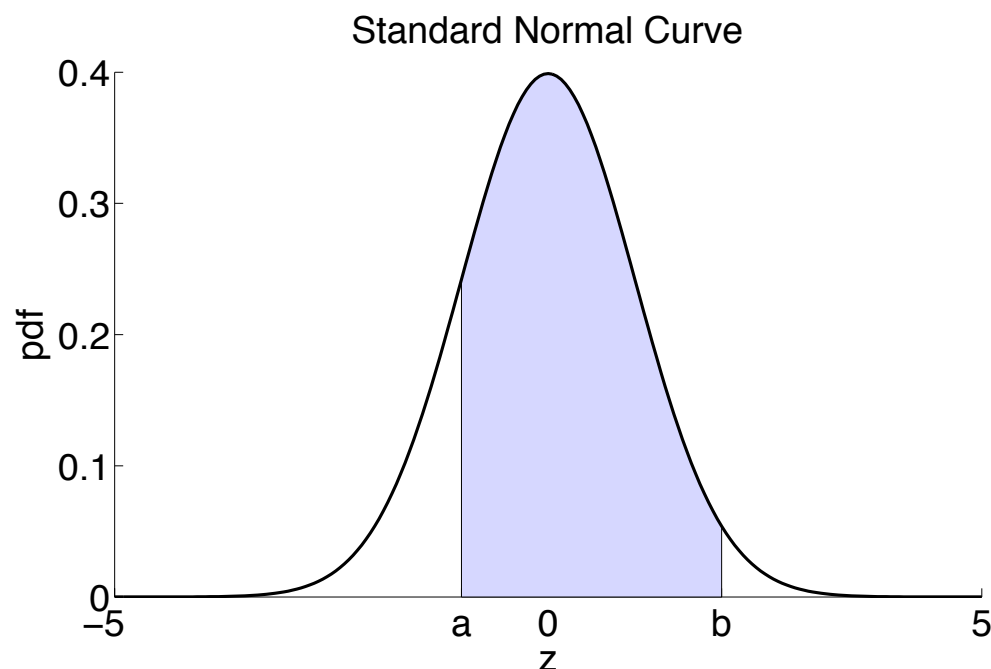
$$\Phi(1.96) \approx 0.9750 \quad \Phi^{-1}(.9750) \approx 1.96$$

Matlab: `normcdf(1.96)` `norminv(.9750)`

R: `pnorm(1.96)` `qnorm(.9750)`

- We will see shortly how to convert between an arbitrary normal distribution (any μ, σ) and the standard normal distribution.
- The commands above allow additional arguments to specify μ and σ , e.g., `normcdf(1.96, 0, 1)`.
- R also can work with the right tail directly:
`pnorm(1.96, lower.tail = FALSE) \approx 0.9750`
`qnorm(0.9750, lower.tail = FALSE) \approx 1.96`

Standard normal distribution — areas



- The area between $z = a$ and $z = b$ is

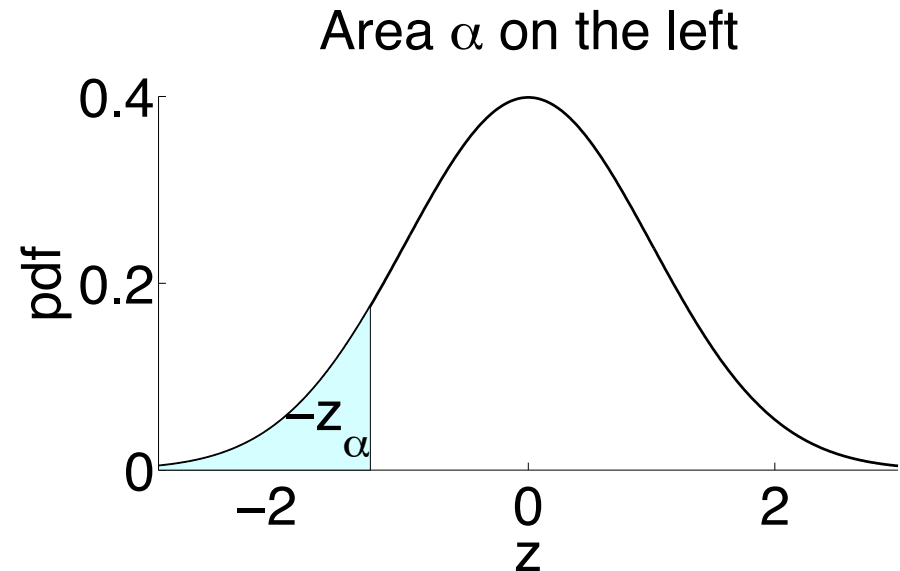
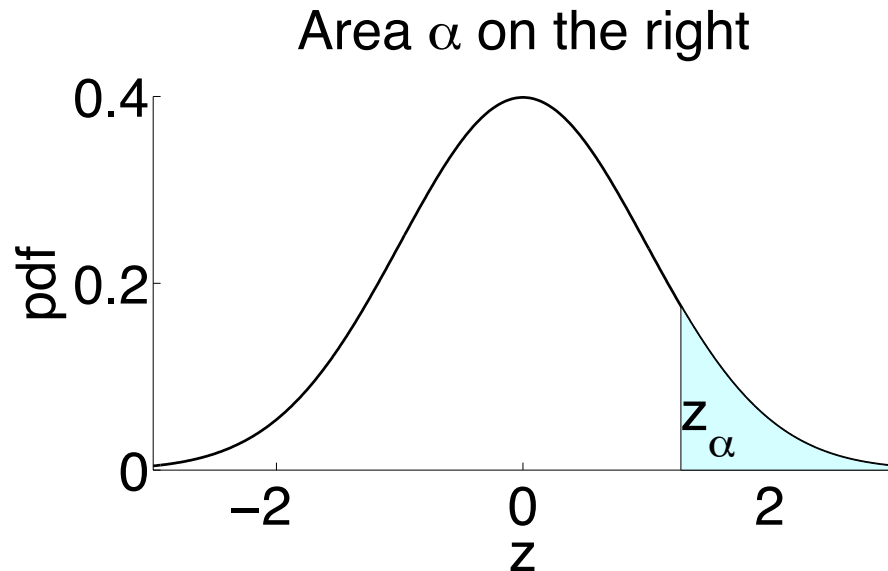
$$P(a \leq Z \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-t^2/2} dt = \Phi(b) - \Phi(a)$$

- $P(1.51 \leq Z \leq 1.62) = \Phi(1.62) - \Phi(1.51) = 0.9474 - 0.9345 = 0.0129$

Matlab: `normcdf(1.62) - normcdf(1.51)`

R: `pnorm(1.62) - pnorm(1.51)`

Standard normal distribution — symmetries of areas



- Area right of z is $P(Z > z) = 1 - \Phi(z)$.
- By symmetry, the area left of $-z$ and the area right of z are equal:

$$\Phi(-z) = 1 - \Phi(z)$$

$$\Phi(-1.51) = 1 - \Phi(1.51) = 1 - 0.9345 = 0.0655$$

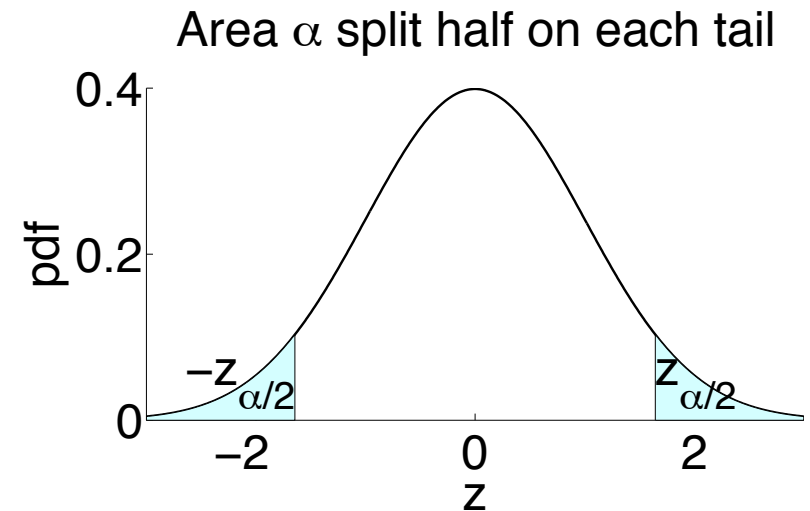
- Area between $z = \pm a$:

$$\Phi(a) - \Phi(-a) = \Phi(a) - (1 - \Phi(a)) = 2\Phi(a) - 1$$

$$\Phi(1.51) - \Phi(-1.51) = 2\Phi(1.51) - 1 \approx .8690$$

Central area

- Area between $z = \pm 1$ is $\approx 68.27\%$.
- Area between $z = \pm 2$ is $\approx 95.45\%$.
- Area between $z = \pm 3$ is $\approx 99.73\%$.



Find the center part containing 95% of the area

- Put 2.5% of the area at the upper tail, 2.5% at the lower tail, and 95% in the middle.
- The value of z putting 2.5% at the top gives $\Phi(z) = 1 - 0.025 = 0.975$.
- **Notation:** $z_{.025} = 1.96$. The area between $z = \pm 1.96$ is about 95%.
- For 99% in the middle, 0.5% on each side, use $z_{.005} \approx 2.58$.

Areas on normal curve for arbitrary μ, σ

$$P(a \leq X \leq b) = \int_a^b \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) dx$$

- Substitute $z = \frac{x - \mu}{\sigma}$ (or $x = \sigma z + \mu$) into the x integral to turn it into the standard normal integral:

$$\begin{aligned} P\left(\frac{a - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{b - \mu}{\sigma}\right) &= P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right) \\ &= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right) \end{aligned}$$

- The ***z-score*** of x is $z = \frac{x - \mu}{\sigma}$.

Binomial distribution

Compute $P(43 \leq X \leq 51)$ when $n = 60, p = 3/4$

Binomial: $n = 60, p = 3/4$

k	$P(X = k) = \binom{60}{k} (.75)^k (.25)^{60-k}$
43	0.09562
44	0.11083
45	0.11822
46	0.11565
47	0.10335
48	0.08397
49	0.06169
50	0.04071
51	0.02395
Total	0.75404

- **Mean**

$$\mu = np = 60(3/4) = 45$$

- **Standard deviation**

$$\begin{aligned}\sigma &= \sqrt{np(1-p)} \\ &= \sqrt{60(3/4)(1/4)} \\ &= \sqrt{11.25} \approx 3.354101966\end{aligned}$$

- **Mode (k with max pdf)**

$$\begin{aligned}&\lfloor np + p \rfloor \\ &= \lfloor 60(3/4) + (3/4) \rfloor \\ &= \lfloor 45\frac{3}{4} \rfloor = 45\end{aligned}$$

Mode of a distribution

The **mode** of random variable X is the value k at which the pdf is maximum.

Mode of binomial distribution when $0 < p < 1$

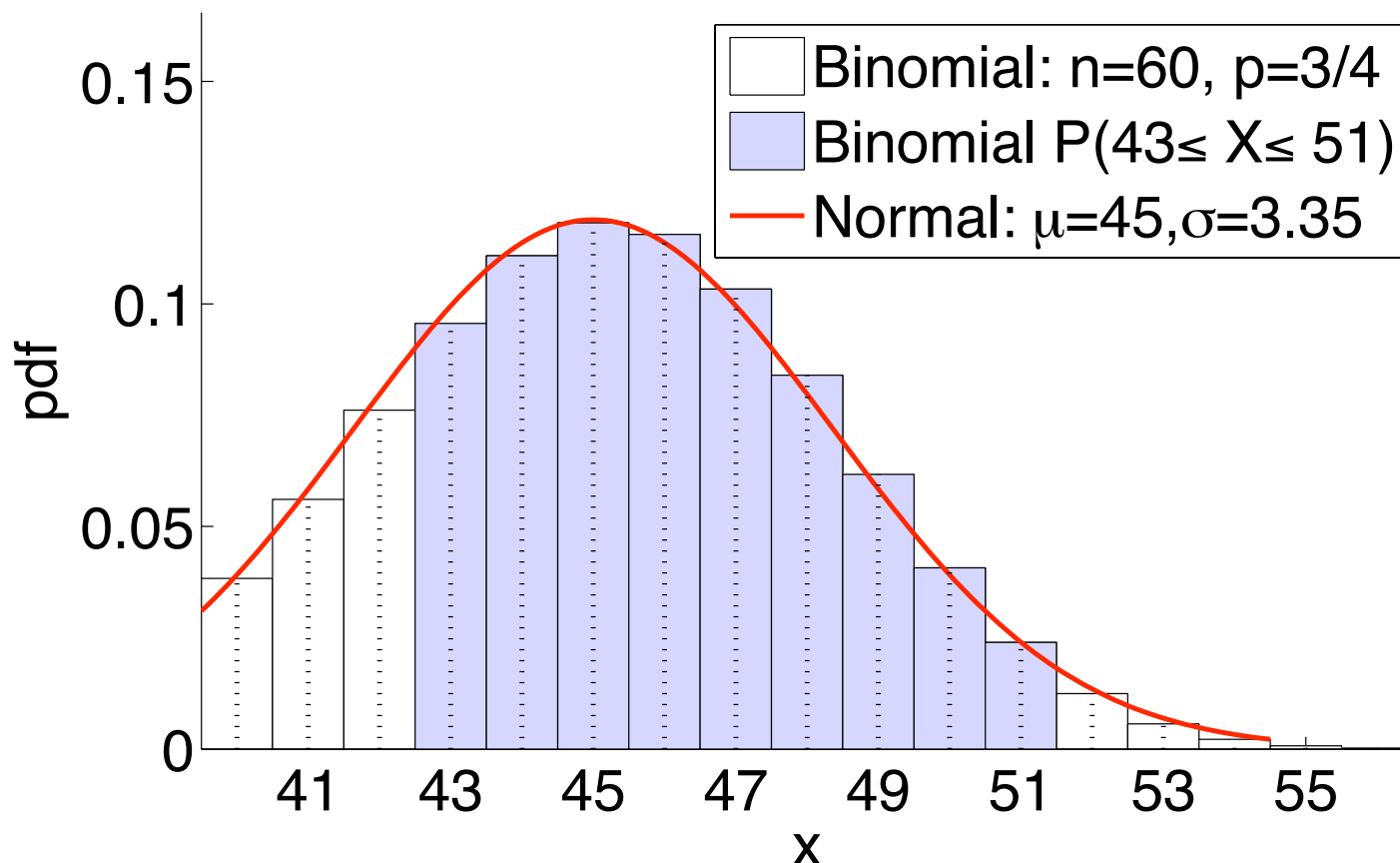
- The mode is $\lfloor (n + 1)p \rfloor$.
- *Exception:* If $(n + 1)p$ is an integer then $(n + 1)p$ and $(n + 1)p - 1$ are tied as the mode.
- The mode is within 1 of the mean np .
- When np is an integer, the mode equals the mean.

Binomial and normal distributions

Binomial

k	$P(X = k)$
43	0.09562
44	0.111083
45	0.11822
46	0.11565
47	0.10335
48	0.08397
49	0.06169
50	0.04071
51	0.02395
Total	0.75404

Normal approximation to binomial



$P(X = k)$ is shown as a rectangle centered above $X = k$:

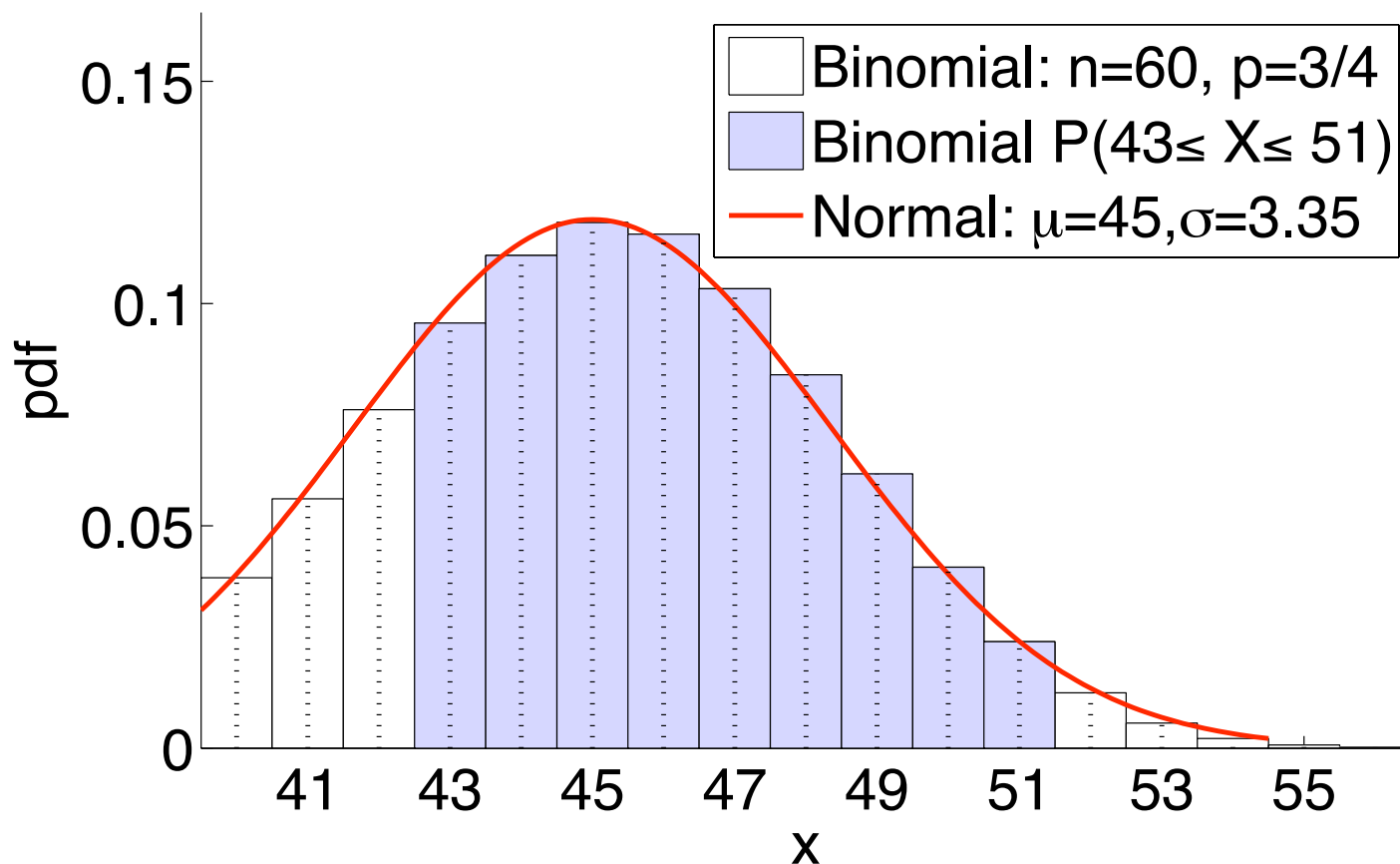
- Height $P(X = k)$.
- Extent $k \pm 1/2$ gives width 1.
- Area $1 \cdot P(X = k) = P(X = k)$.
- Area of all purple rectangles is $P(43 \leq X \leq 51)$.

Binomial and normal distributions

Binomial

k	$P(X = k)$
43	0.09562
44	0.111083
45	0.11822
46	0.11565
47	0.10335
48	0.08397
49	0.06169
50	0.04071
51	0.02395
Total	0.75404

Normal approximation to binomial



- The binomial distribution is only defined at the integers, and is very close to the normal distribution shown.
- We will approximate the probability $P(43 \leq X \leq 51)$ we had above by the corresponding one for the normal distribution.
- Riemann sums in Calculus: area under curve \approx area of rectangles

Normal approximation to binomial, step 1

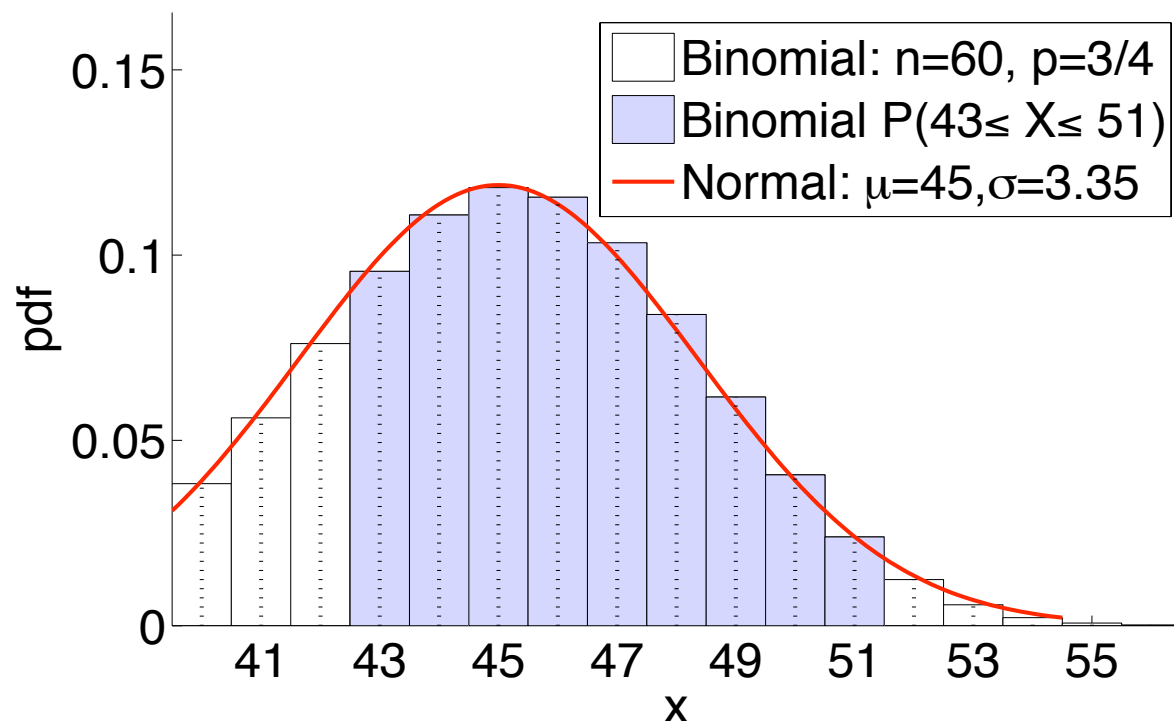
Compute corresponding parameters

- We want to approximate $P(a \leq X \leq b)$ in a binomial distribution. We'll use $n = 60$, $p = 3/4$ and approximate $P(43 \leq X \leq 51)$.
- **Determine μ , σ :**
$$\mu = np = 60(3/4) = 45$$
$$\sigma = \sqrt{np(1-p)} = \sqrt{11.25} \approx 3.354$$
- The normal distribution with those same values of μ , σ is a good approximation to the binomial distribution *provided* $\mu \pm 3\sigma$ are both between 0 and n .
- **Check:**
$$\mu - 3\sigma \approx 45 - 3(3.354) = 34.938$$
$$\mu + 3\sigma \approx 45 + 3(3.354) = 55.062$$
are both between 0 and 60, so we may proceed.
- **Note:** Some applications are more strict and may require $\mu \pm 5\sigma$ or more to be between 0 and n . Since $\mu + 5\sigma \approx 61.771$, this would fail at that level of strictness.

Normal approximation to binomial, step 2

Continuity correction

Normal approximation to binomial



- The binomial distribution is discrete ($X = \text{integers}$) but the normal distribution is continuous.

- The sum $P(X = 43) + \dots + P(X = 51)$ has 9 terms, corresponding to the area of the 9 rectangles in the picture.
- The area under the normal distribution curve from $42.5 \leq X \leq 51.5$ approximates the area of those rectangles.
- Change binomial $P(43 \leq X \leq 51)$ to normal $P(42.5 \leq X \leq 51.5)$.

Normal approximation to binomial, steps 3–4

3. Convert to z -scores

4. Use the normal distribution to approximately evaluate it

- For random variable X with mean μ and standard deviation σ ,

- The z -score of a value x is $z = \frac{x - E(X)}{SD(X)} = \frac{x - \mu}{\sigma}$.

- The random variable Z is $Z = \frac{X - E(X)}{SD(X)} = \frac{X - \mu}{\sigma}$.

- Convert to z -scores:

$$\begin{aligned} P(42.5 \leq X \leq 51.5) &= P\left(\frac{42.5 - 45}{\sqrt{11.25}} \leq \frac{X - 45}{\sqrt{11.25}} \leq \frac{51.5 - 45}{\sqrt{11.25}}\right) \\ &= P(-.7453559926 \leq Z \leq 1.937925581) \end{aligned}$$

- Approximate this by the standard normal distribution cdf:

$$\approx \Phi(1.937925581) - \Phi(-.7453559926)$$

$$\approx 0.7456555785$$

- This is close to the true answer (apart from rounding errors)

$$P(43 \leq X \leq 51) = 0.75404 \text{ we got from the binomial distribution.}$$

Estimating fraction of successes instead of number of successes

- What is the value of p in the binomial distribution?
- Estimate it: flip a coin n times and divide the # heads by n .
- Let X = binomial distribution for n flips, probability p of heads.
- Let $\bar{X} = X/n$ be the fraction of flips that are heads.
- \bar{X} is discrete, with possible values $\frac{0}{n}, \frac{1}{n}, \frac{2}{n}, \dots, \frac{n}{n}$.
- $P(\bar{X} = \frac{k}{n}) = P(X = k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k} & \text{for } k = 0, 1, \dots, n; \\ 0 & \text{otherwise.} \end{cases}$
- **Mean** $E(\bar{X}) = E(X/n) = E(X)/n = np/n = p$.
- **Variance** $\text{Var}(\bar{X}) = \text{Var}\left(\frac{X}{n}\right) = \frac{\text{Var}(X)}{n^2} = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$.
- **Standard deviation** $\text{SD}(\bar{X}) = \sqrt{p(1-p)/n}$.

Normal approximation for fraction of successes

- n flips, probability p of heads, \bar{X} =observed fraction of heads

Mean $E(\bar{X}) = p$

Variance $\text{Var}(\bar{X}) = p(1 - p)/n$

Standard deviation $\text{SD}(\bar{X}) = \sqrt{p(1 - p)/n}$

- The Z transformation of \bar{X} is

$$Z = \frac{\bar{X} - E(\bar{X})}{\text{SD}(\bar{X})} = \frac{\bar{X} - p}{\sqrt{p(1 - p)/n}}$$

and value $\bar{X} = \bar{x}$ has z -score $z = \frac{\bar{x} - p}{\sqrt{p(1 - p)/n}}$.

- For k heads in n flips,

- The z -score of $X = k$ is $z_1 = \frac{k - np}{\sqrt{np(1 - p)}}$.

- The z -score of $\bar{X} = k/n$ is $z_2 = \frac{(k/n) - p}{\sqrt{p(1 - p)/n}}$.

- These are equal! Divide the numerator and denominator of z_1 by n to get z_2 .

Normal approximation for fraction of successes

- For $n = 60$ flips of a coin with $p = \frac{3}{4}$, we'll estimate $P\left(\frac{43}{60} \leq \bar{X} \leq \frac{51}{60}\right)$.
- The exact answer equals $P(43 \leq X \leq 51) \approx 0.75404$.

- **Step 1: Determine mean and SD**

$$E(\bar{X}) = p = .75$$

$$SD(\bar{X}) = \sqrt{p(1-p)/n} = \sqrt{(.75)(.25)/60} = \sqrt{0.003125} \approx 0.05590$$

- **Verify approximation is valid: Mean \pm 3 SD between 0 and 1**

$$\text{Mean} - 3 \text{ SD} = 0.58229$$

$$\text{Mean} + 3 \text{ SD} = 0.91770$$

Both are between 0 and 1.

- **Step 2: Continuity correction**

$$P\left(\frac{43}{60} \leq \bar{X} \leq \frac{51}{60}\right) = P\left(\frac{42.5}{60} \leq \bar{X} \leq \frac{51.5}{60}\right)$$

- **Step 3: z-scores**

- **Step 4: Evaluate approximate answer using normal distribution**

Normal approximation for fraction of successes

$$\begin{aligned}P\left(\frac{43}{60} \leq \bar{X} \leq \frac{51}{60}\right) &= P\left(\frac{42.5}{60} \leq \bar{X} \leq \frac{51.5}{60}\right) \\&= P(0.70833 \leq \bar{X} \leq .85833) \\&= P\left(\frac{0.70833 - E(\bar{X})}{SD(\bar{X})} \leq \frac{\bar{X} - E(\bar{X})}{SD(\bar{X})} \leq \frac{.85833 - E(\bar{X})}{SD(\bar{X})}\right) \\&= P\left(\frac{0.70833 - .75}{0.05590} \leq Z \leq \frac{.85833 - .75}{0.05590}\right) \\&= P(-.74535 \leq Z \leq 1.93792) \\&= \Phi(1.93792) - \Phi(-.74535) \approx 0.74565\end{aligned}$$

Mean and SD of sums and averages of i.i.d. random variables

- Let X_1, \dots, X_n be n i.i.d. (independent identically distributed) random variables, each with mean μ and standard deviation σ .
- Let $S_n = X_1 + \dots + X_n$ be their sum and $\bar{X}_n = (X_1 + \dots + X_n)/n = S_n/n$ be their average.

- **Means:**

Sum: $E(S_n) = E(X_1) + \dots + E(X_n) = n E(X_1) = n\mu$

Avg: $E(\bar{X}_n) = E(S_n/n) = n\mu/n = \mu$

- **Variance:**

Sum: $\text{Var}(S_n) = \text{Var}(X_1) + \dots + \text{Var}(X_n) = n \text{Var}(X_1) = n\sigma^2$

Avg: $\text{Var}(\bar{X}_n) = \text{Var}(S_n)/n^2 = n\sigma^2/n^2 = \sigma^2/n$

- **Standard deviation:**

Sum: $\text{SD}(S_n) = \sigma \sqrt{n}$

Avg: $\text{SD}(\bar{X}_n) = \sigma / \sqrt{n}$

Terminology for different types of standard deviation

- The *standard deviation* (SD) of a trial (each X_i) is σ
- The *standard error* (SE) of the sum is $\sigma \sqrt{n}$
- The *standard error* (SE) of the average is σ / \sqrt{n}

Z-scores of sums and averages

	For sum S_n	For average \bar{X}_n
Mean:	$E(S_n) = n\mu$	$E(\bar{X}_n) = \mu$
Variance:	$\text{Var}(S_n) = n\sigma^2$	$\text{Var}(\bar{X}_n) = \sigma^2/n$
Standard Deviation:	$\text{SD}(S_n) = \sigma\sqrt{n}$	$\text{SD}(\bar{X}_n) = \sigma/\sqrt{n}$
Z-scores:	$Z = \frac{S_n - E(S_n)}{\text{SD}(S_n)} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$	$Z = \frac{\bar{X}_n - E(\bar{X}_n)}{\text{SD}(\bar{X}_n)} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$

Z-scores of sum and average are equal! Divide the numerator and denominator of Z of the sum by n to get Z of the average.

$$Z_{\text{sum}} = \frac{(S_n - n\mu)/n}{(\sigma\sqrt{n})/n} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = Z_{\text{avg}}$$

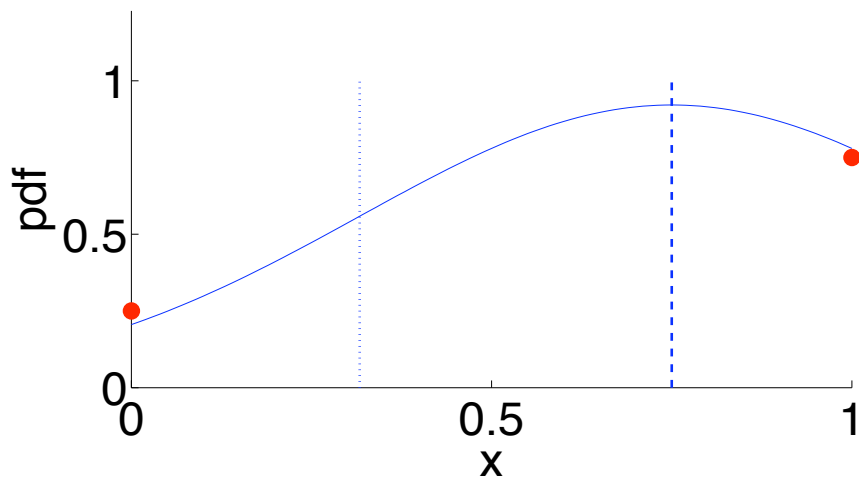
Theorem (Central Limit Theorem — abbreviated CLT)

For n i.i.d. random variables X_1, \dots, X_n with sum $S_n = X_1 + \dots + X_n$ and average $\bar{X}_n = S_n/n$, and any real numbers $a < b$,

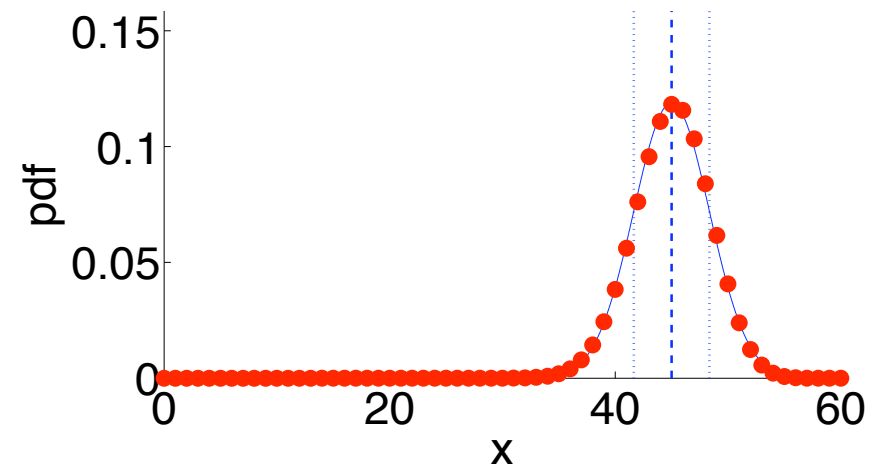
$$P\left(a \leq \frac{S_n - n\mu}{\sigma\sqrt{n}} \leq b\right) = P\left(a \leq \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq b\right) \approx \Phi(b) - \Phi(a)$$

if n is large enough. As $n \rightarrow \infty$, the approximation becomes exact equality.

Binomial $n=1, p=0.75; \mu=0.75, \sigma=0.43$



Binomial $n=60, p=0.75; \mu=45.00, \sigma=3.35$



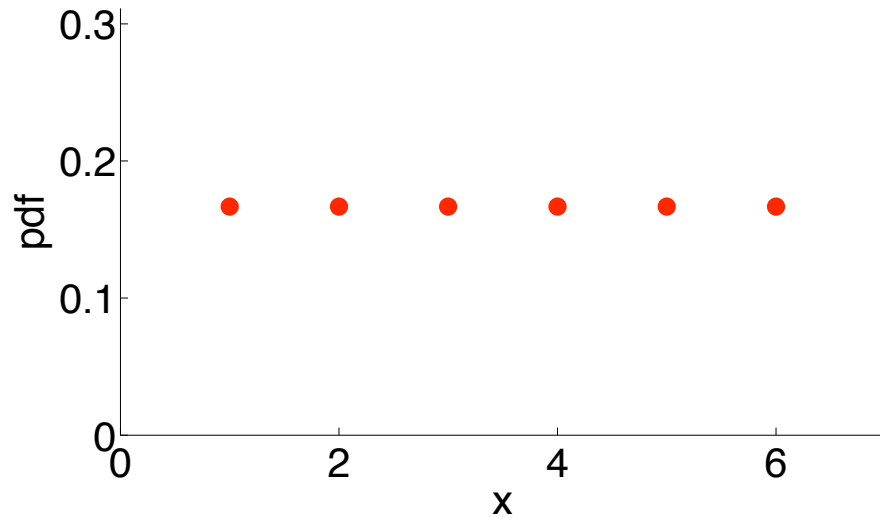
Interpretation of Central Limit Theorem

- As n increases, the pdf more and more closely resembles a normal distribution.
- However, the pdf is defined as 0 in-between the red points shown, if it's a discrete distribution.
- The cdfs are approximately equal everywhere on the continuum.
- Probabilities of intervals for sums or averages of enough i.i.d. variables can be approximately evaluated using the normal distribution.

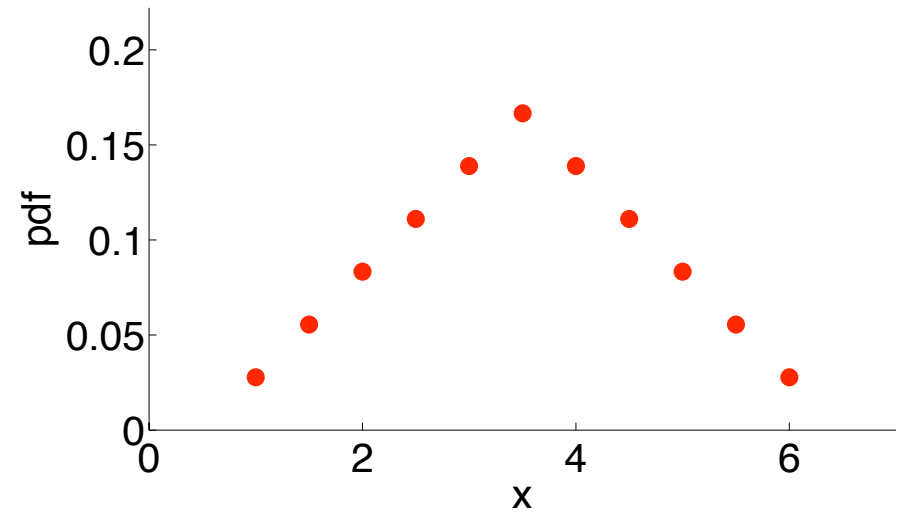
Repeated rolls of a die

One roll: $\mu = 3.5$, $\sigma = \sqrt{35/12} \approx 1.71$

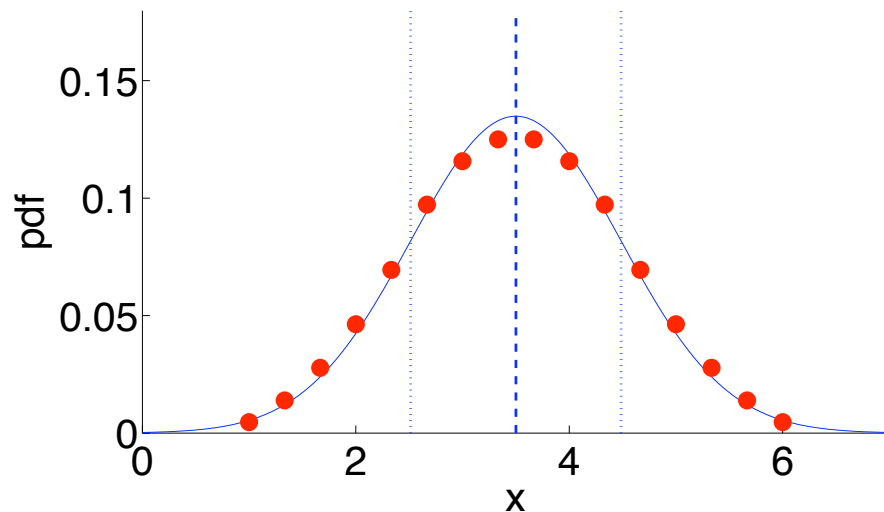
Average of 1 roll of die; $\mu=3.50$, $\sigma=1.71$



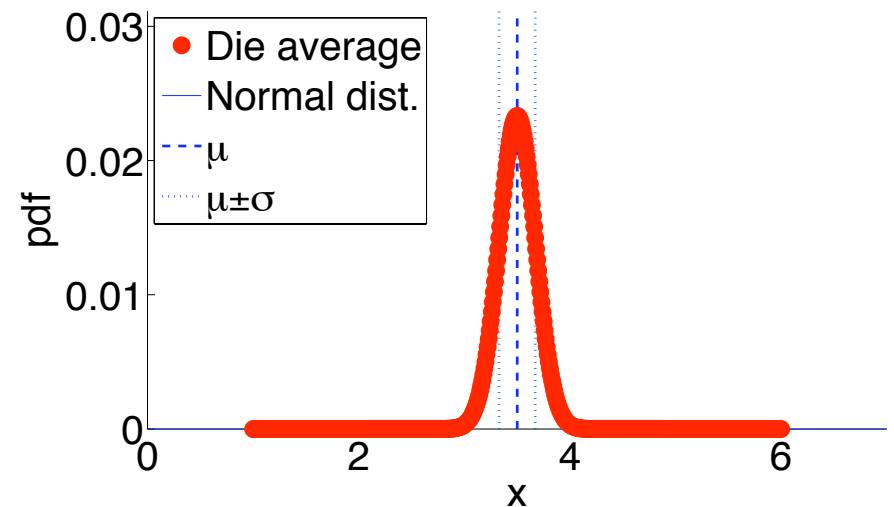
Average of 2 rolls of die; $\mu=3.50$, $\sigma=1.21$



Average of 3 rolls of die; $\mu=3.50$, $\sigma=0.99$



Average of 100 rolls of die; $\mu=3.50$, $\sigma=0.17$



Repeated rolls of a die

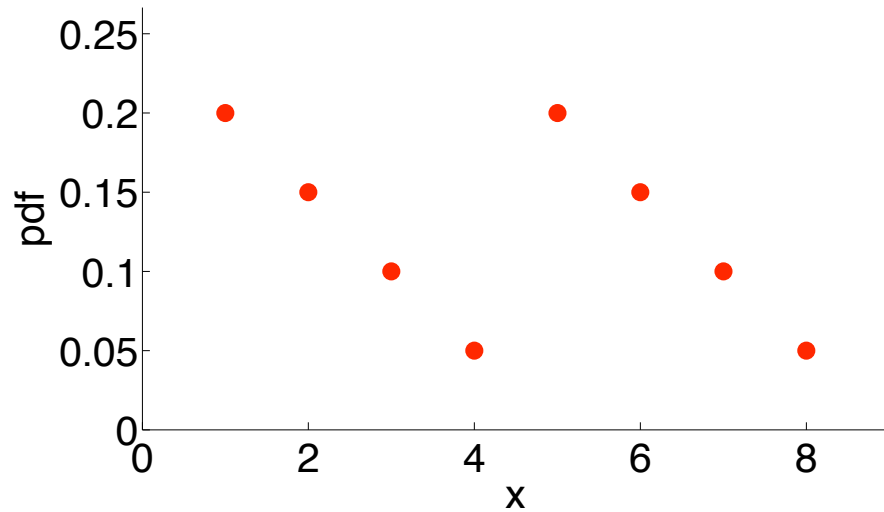
Find n so that at least 95% of the time, the average of n rolls of a die is between 3 and 4.

- $P(3 \leq \bar{X} \leq 4) = P\left(\frac{3-\mu}{\sigma/\sqrt{n}} \leq \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq \frac{4-\mu}{\sigma/\sqrt{n}}\right)$
- Plug in $\mu = 3.5$ and $\sigma = \sqrt{35/12}$.
- $P(3 \leq \bar{X} \leq 4) = P\left(-\frac{1/2}{\sqrt{35/(12n)}} \leq Z \leq \frac{1/2}{\sqrt{35/(12n)}}\right)$
- Recall the center 95% of the area on the standard normal curve is between $z = \pm 1.96$.
- $\frac{1/2}{\sqrt{35/(12n)}} \geq 1.96 \quad \Rightarrow \quad n \geq (1.96)^2 \frac{35/12}{(1/2)^2} \approx 44.81$
- n is an integer so $n \geq 45$

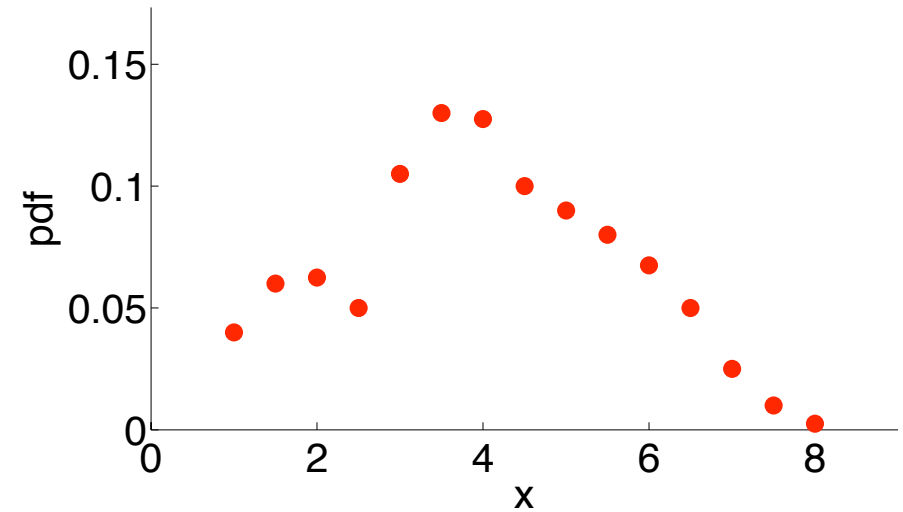
“Sawtooth” distribution (made up as demo)

One trial: $\mu = 4, \sigma \approx 2.24$

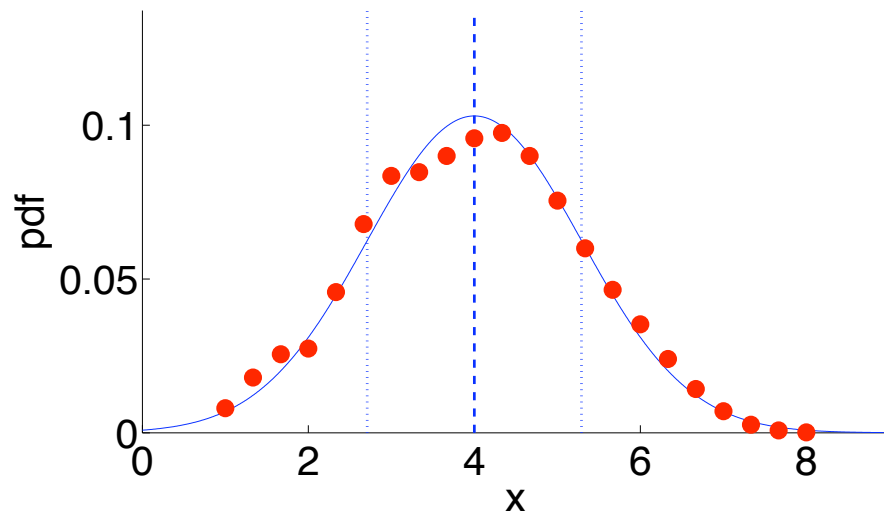
Average of 1 trial; $\mu=4.00, \sigma=2.24$



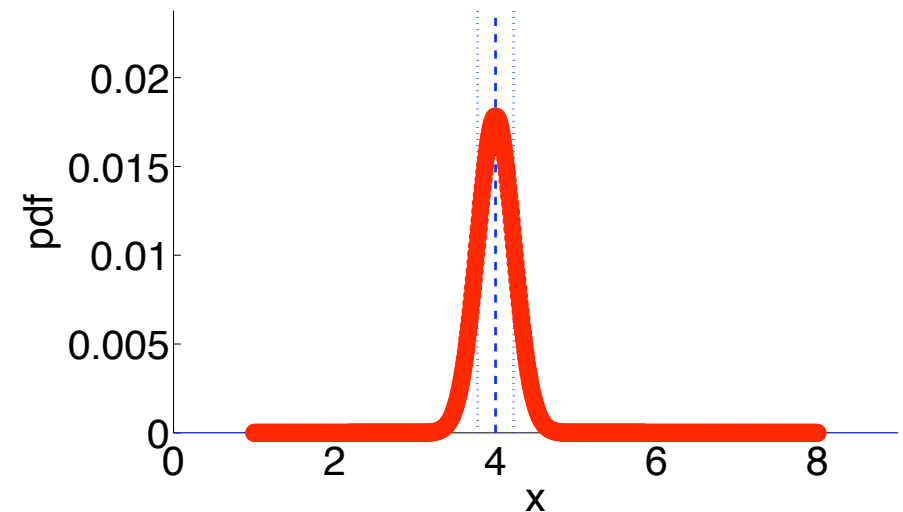
Average of 2 trials; $\mu=4.00, \sigma=1.58$



Average of 3 trials; $\mu=4.00, \sigma=1.29$



Average of 100 trials; $\mu=4.00, \sigma=0.22$



Binomial distribution (n, p)

- A *Bernoulli trial* is to flip a coin once and count the number of heads,

$$X_1 = \begin{cases} 1 & \text{probability } p; \\ 0 & \text{probability } 1 - p. \end{cases}$$

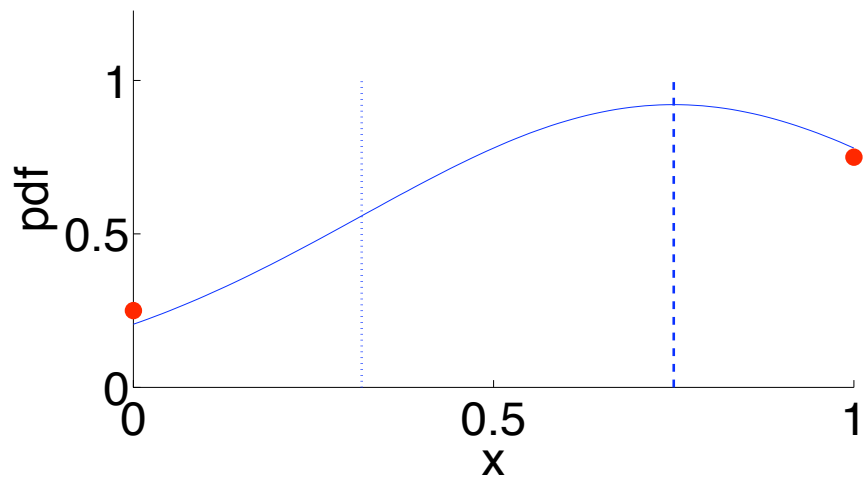
Mean $E(X_1) = p$, standard deviation $SD(X_1) = \sqrt{p(1-p)}$.

- The binomial distribution is the sum of n i.i.d. Bernoulli trials.
Mean $\mu = np$, standard deviation $\sigma = \sqrt{np(1-p)}$.
- The binomial distribution is approximated pretty well by the normal distribution when $\mu \pm 3\sigma$ are between 0 and n .

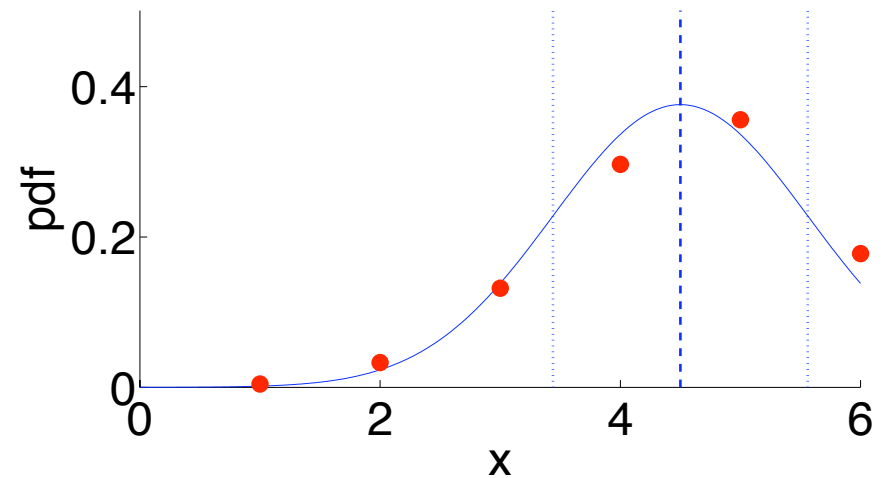
Binomial distribution (n, p)

One flip: $\mu = p = .75$, $\sigma = \sqrt{p(1-p)} = \sqrt{.1875} \approx 0.4330$

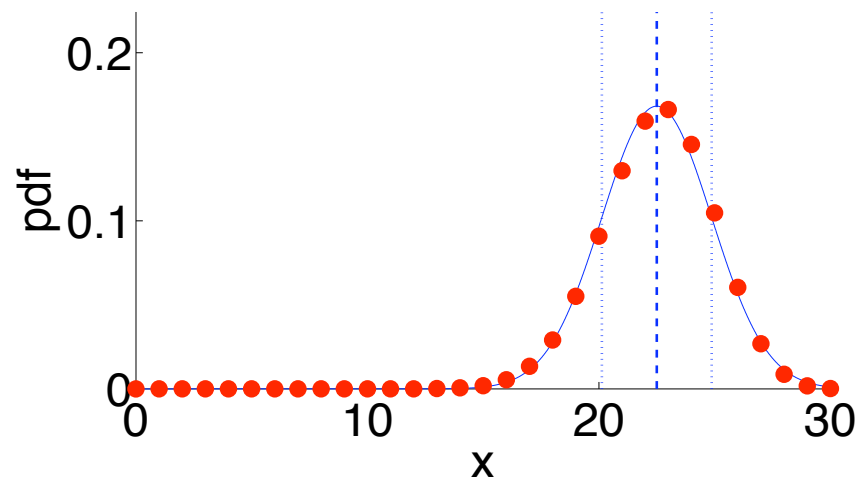
Binomial $n=1, p=0.75; \mu=0.75, \sigma=0.43$



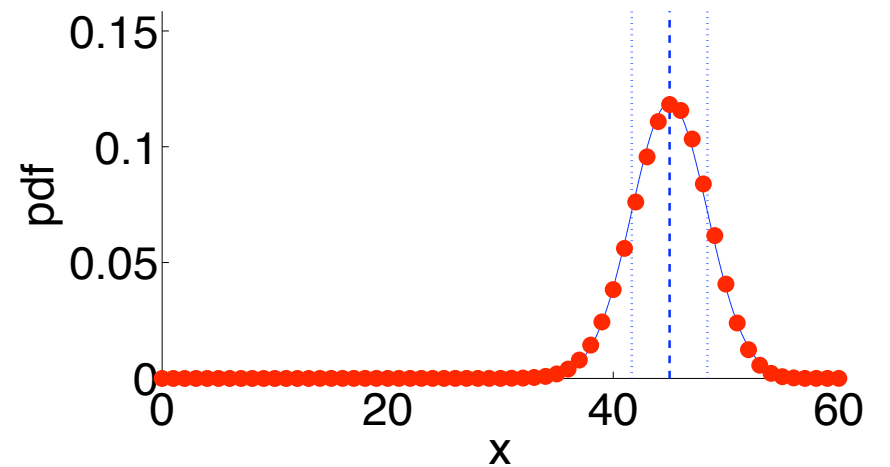
Binomial $n=6, p=0.75; \mu=4.50, \sigma=1.06$



Binomial $n=30, p=0.75; \mu=22.50, \sigma=2.37$



Binomial $n=60, p=0.75; \mu=45.00, \sigma=3.35$

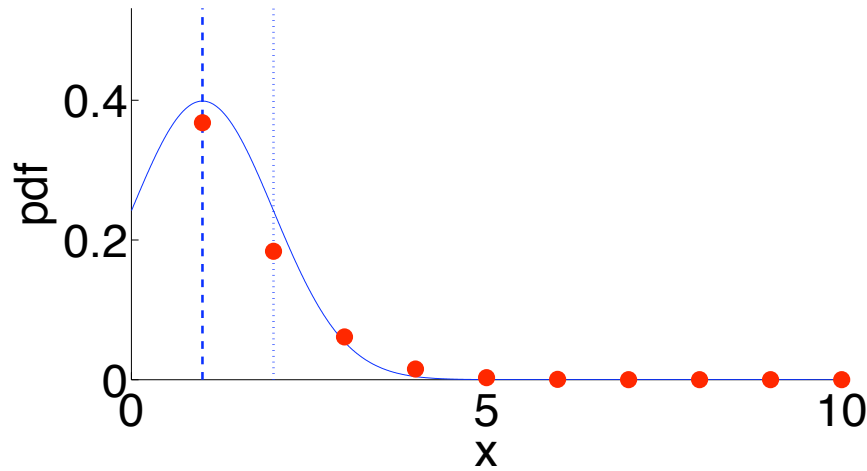


Poisson distribution (μ or $\mu = \lambda d$)

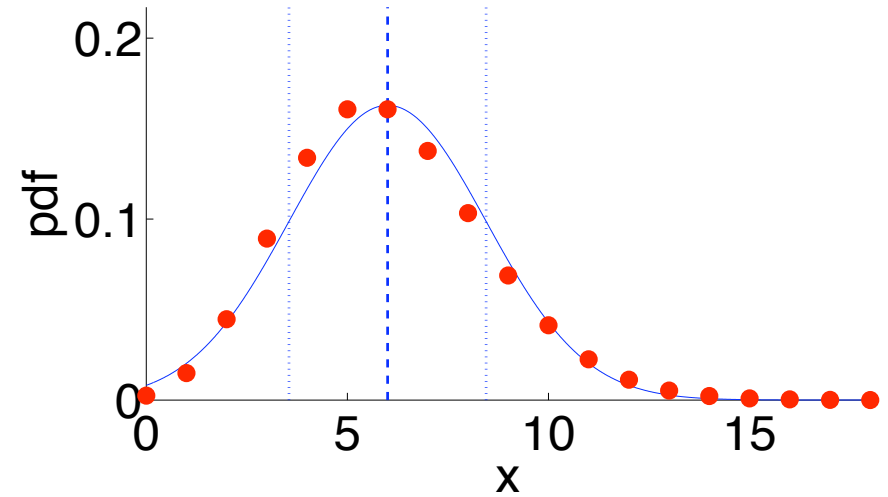
- **Mean:** μ (same as the Poisson parameter)
Standard deviation: $\sigma = \sqrt{\mu}$.
- It is approximated pretty well by the normal distribution when $\mu \geq 5$.
- The reason the Central Limit Theorem applies is that a Poisson distribution with parameter μ equals the sum of n i.i.d. Poissons with parameter μ/n .
- The Poisson distribution has infinite range $x = 0, 1, 2, \dots$ and the normal distribution has infinite range $-\infty < x < \infty$ (reals). Both are truncated in the plots.

Poisson distribution (μ)

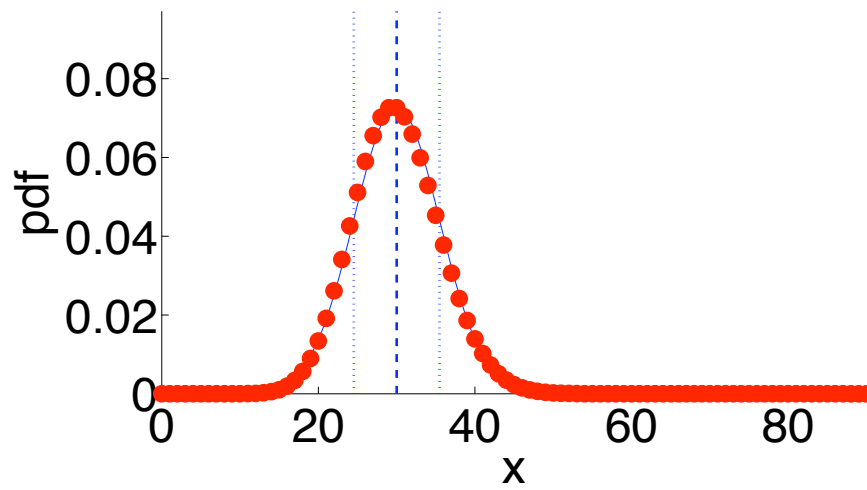
Poisson $\mu=1$; $\sigma=1.00$



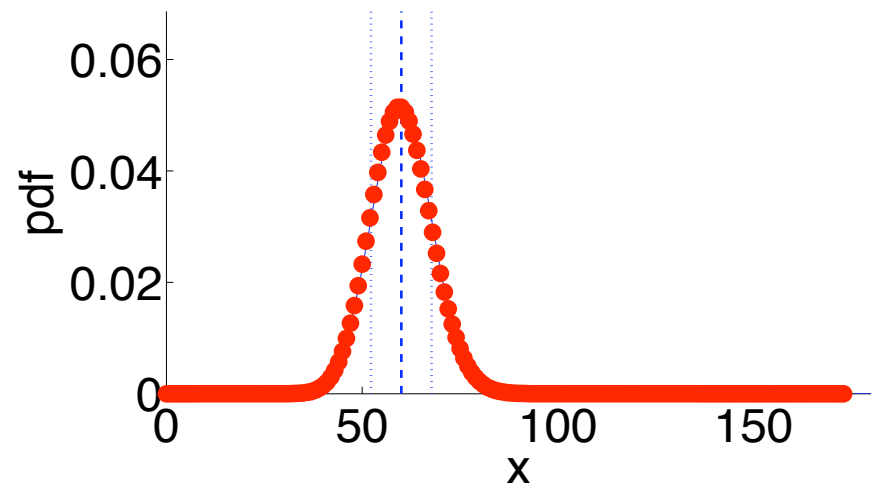
Poisson $\mu=6$; $\sigma=2.45$



Poisson $\mu=30$; $\sigma=5.48$



Poisson $\mu=60$; $\sigma=7.75$



Geometric and negative binomial distributions

Geometric distribution (p)

- X is the number of flips until the first heads,

$$p_X(x) = \begin{cases} (1-p)^{x-1}p & \text{if } x = 1, 2, 3, \dots; \\ 0 & \text{otherwise.} \end{cases}$$

- The pdf plot doesn't resemble the normal distribution at all.
- **Mean:** $\mu = 1/p$ **Standard deviation:** $\sigma = \sqrt{1-p}/p$

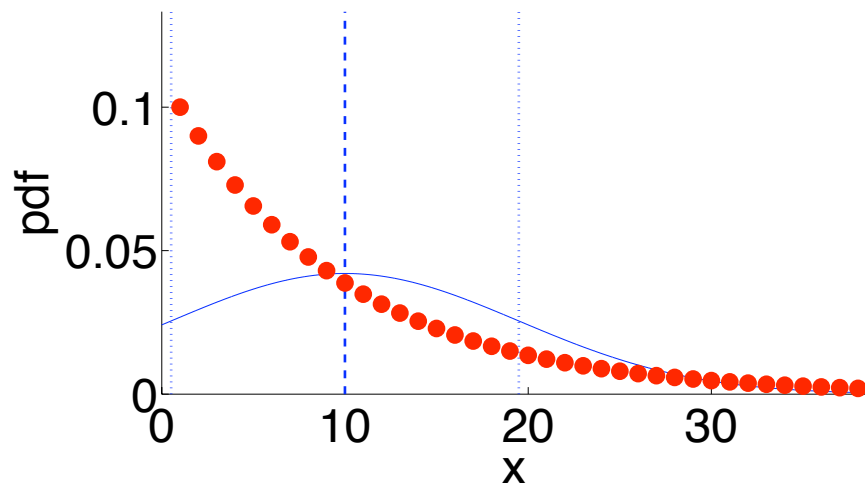
Negative binomial distribution (r, p)

- $r = 1$ is same as geometric distribution.
- $r > 2$: The pdf has a “bell”-like shape, but is not close to the normal distribution unless r is very large.
- **Mean:** $\mu = r/p$ **Standard deviation:** $\sigma = \sqrt{r(1-p)}/p$

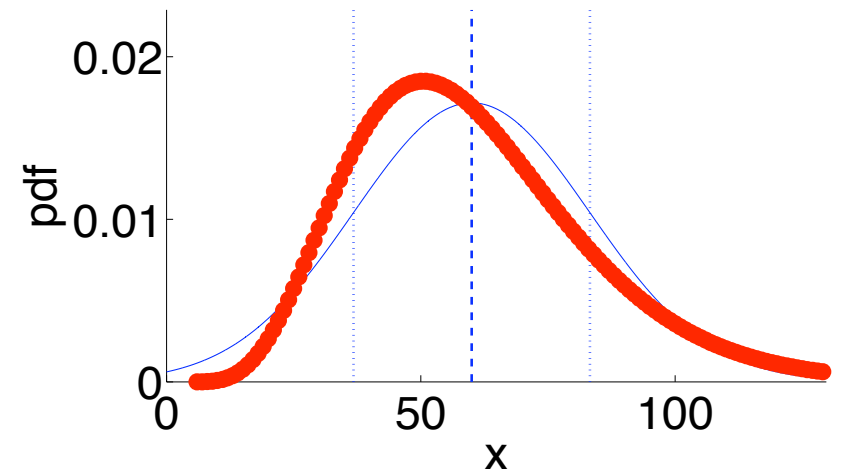
Geometric and negative binomial distributions

Heads with probability $p = .1$

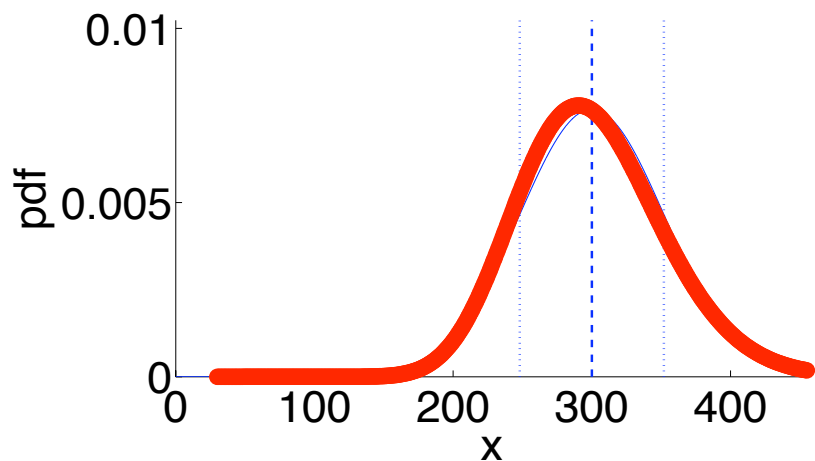
Geometric $p=0.10$; $\mu=10.00$, $\sigma=9.49$



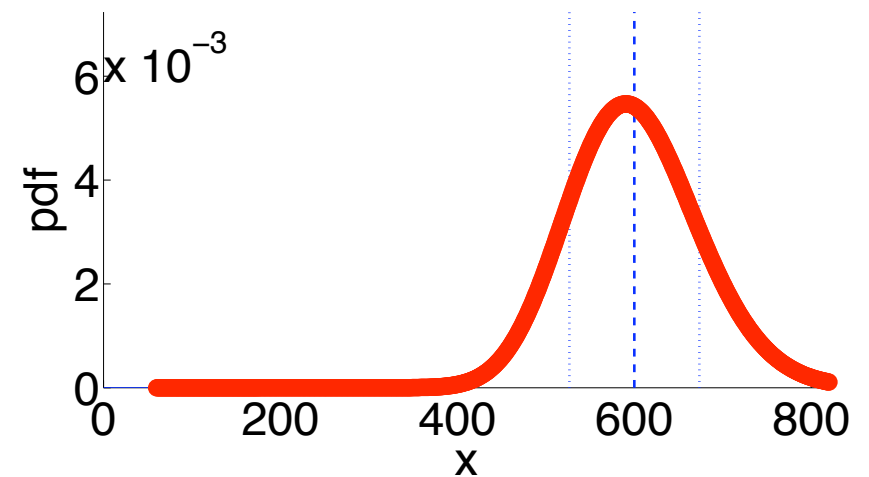
Neg. bin. $r=6$, $p=0.10$; $\mu=60.00$, $\sigma=23.24$



Neg. bin. $r=30$, $p=0.10$; $\mu=300.00$, $\sigma=51.96$



Neg. bin. $r=60$, $p=0.10$; $\mu=600.00$, $\sigma=73.48$



Exponential and gamma distributions

Exponential distribution (λ)

- The exponential distribution doesn't resemble the normal distribution at all.
- **Mean:** $\mu = 1/\lambda$ **Standard deviation:** $\sigma = 1/\lambda$

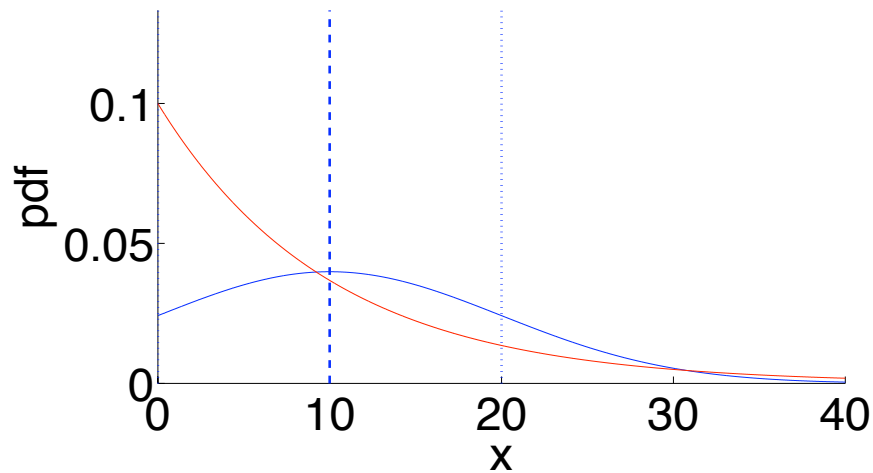
Gamma distribution (r, λ)

- The gamma distribution for $r = 1$ is the exponential distribution.
- The gamma distribution for $r > 1$ does have a “bell”-like shape, but it is not close to the normal distribution until r is very large.
- There is a generalization to allow r to be real numbers, not just integers.
- **Mean:** $\mu = r/\lambda$ **Standard deviation:** $\sigma = \sqrt{r}/\lambda$

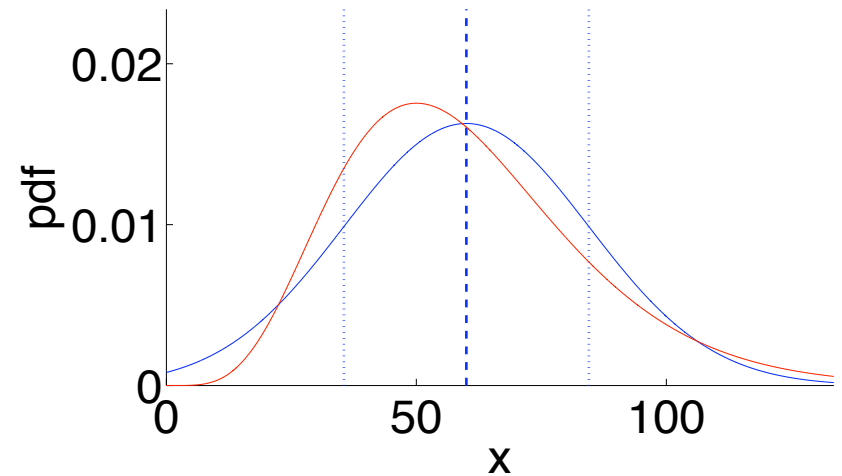
Exponential and gamma distributions

Rate $\lambda = .1$

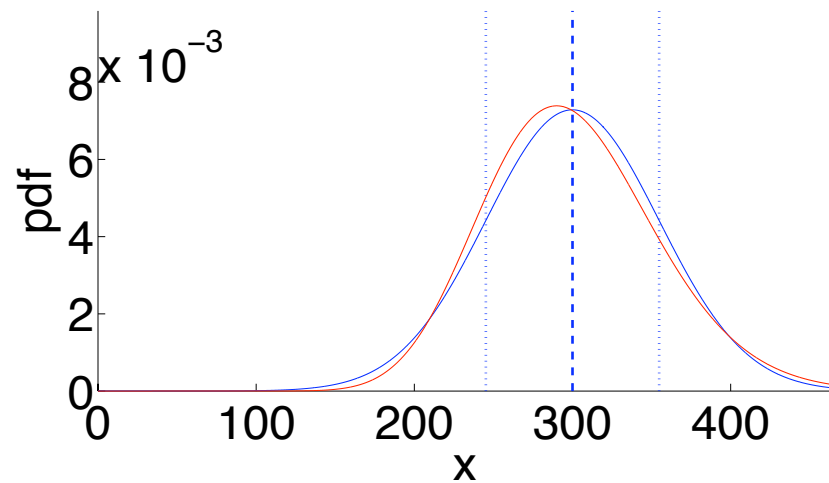
Exponential $\lambda=0.10; \mu=10.00, \sigma=10.00$



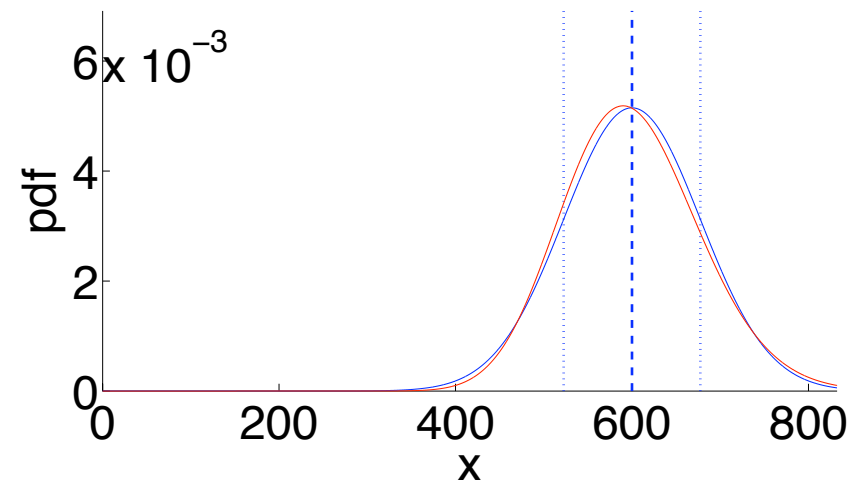
Gamma $r=6, p=0.10; \mu=60.00, \sigma=24.49$



Gamma $r=30, p=0.10; \mu=300.00, \sigma=54.77$



Gamma $r=60, p=0.10; \mu=600.00, \sigma=77.46$



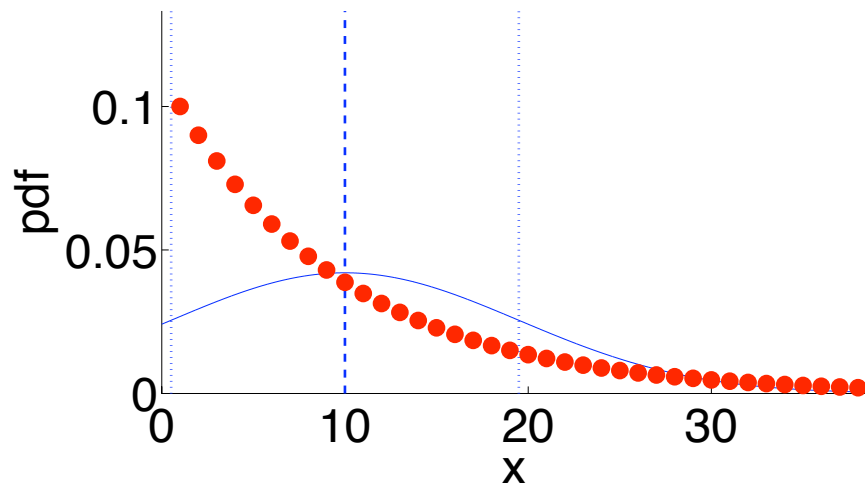
Geometric/Negative binomial vs. Exponential/Gamma

- $p = \lambda$ gives same means for geometric and exponential.
- $p = 1 - e^{-\lambda}$ gives same exponential decay rate for both geometric and exponential distributions.
- $1 - e^{-\lambda} \approx \lambda$ when λ is small.
- This correspondence carries over to the gamma and negative binomial distributions.

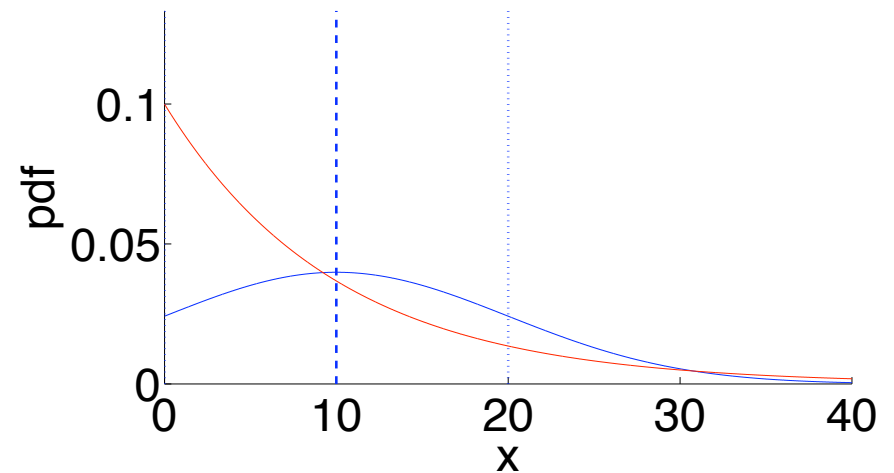
Geometric/negative binomial vs. Exponential/gamma

This is for $p = .1$ vs. $\lambda = .1$; a better fit for $\lambda = .1$ would be $p = 1 - e^{-\lambda} \approx 0.095$

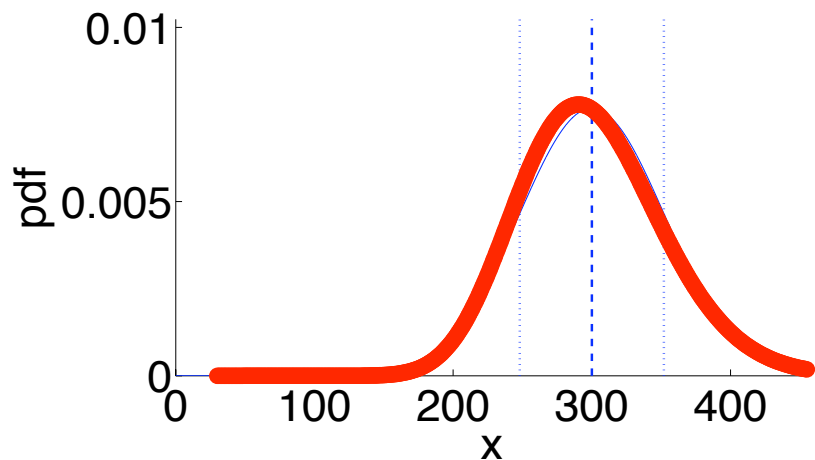
Geometric $p=0.10$; $\mu=10.00$, $\sigma=9.49$



Exponential $\lambda=0.10$; $\mu=10.00$, $\sigma=10.00$



Neg. bin. $r=30$, $p=0.10$; $\mu=300.00$, $\sigma=51.96$



Gamma $r=30$, $p=0.10$; $\mu=300.00$, $\sigma=54.77$

