Chapter 3: Distributions of Random Variables (Part 2)

Days 8-9 topics:
Section 3.3: Normal distribution
Section 3.4: Poisson distribution

3.2. **Normal distribution.**

A random variable X is modeled with a normal distribution if:

- shape: symmetric, unimodal bell curve
- center: mean μ
- spread (variability): standard deviation σ
- Curve is determined by the formula $f(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$, for $-\infty < x < \infty$.

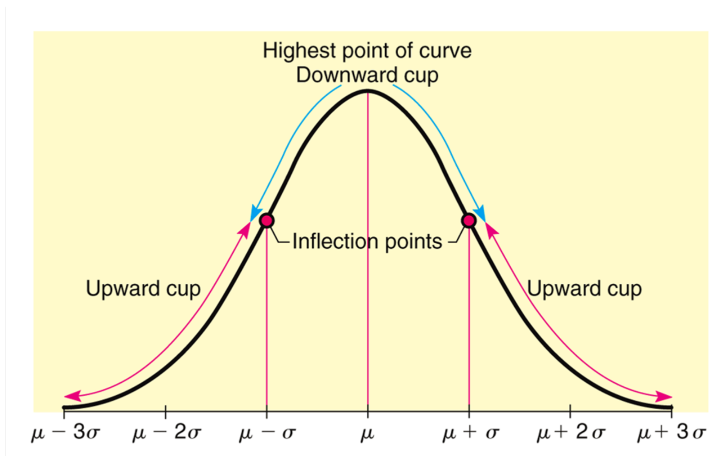


Figure 1. Normal (Gaussian) distribution

$X \sim N(\mu, \sigma)$, Total area =1; L: $\mu = 0, \sigma = 2, 5$, R: $\mu = -1, \sigma = 4$, $\mu = 3, \sigma = 2$

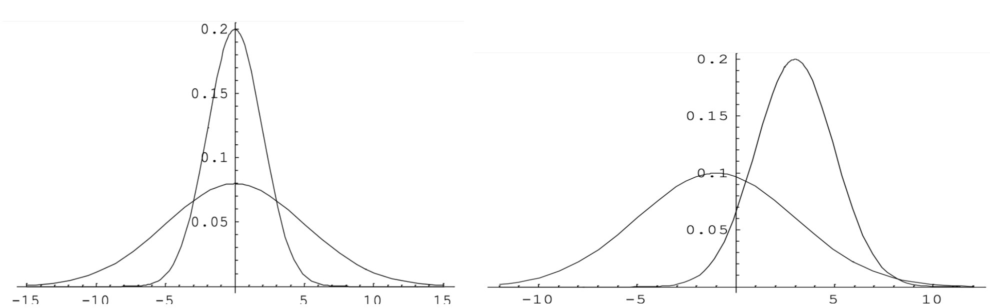


Figure 2. What are the means and standard deviations?

The Empirical Rule (a.k.a. the 68-95-99.7 rule)

For a normal distribution:

- 68% of observations are within one standard deviation (SD) of the mean
- 95% are within two SD's
- 99.7% are within three SD's

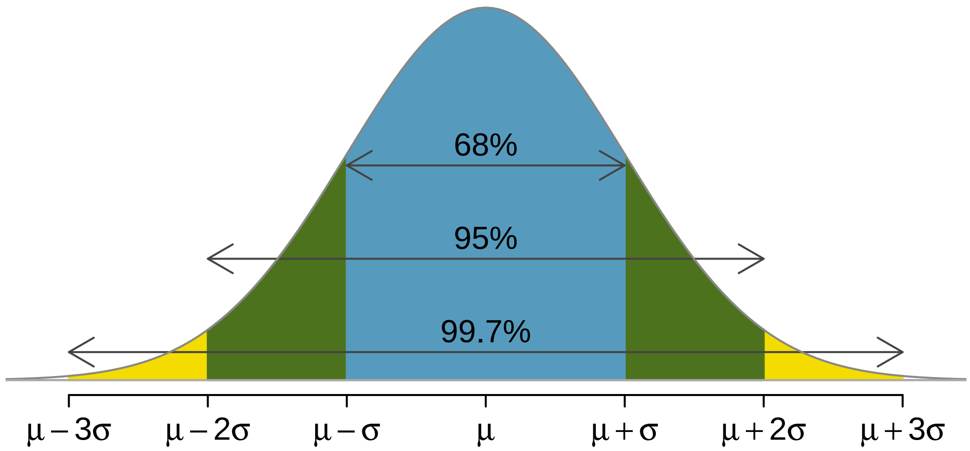


FIGURE 3. Empirical Rule (a.k.a. the 68-95-99.7 rule)

The empirical rule is useful for *estimating* probabilities.

Standard Normal Distribution

- A **standard** normal r.v. has mean 0 and standard deviation 1.
- We denote standard normal r.v.'s with the letter Z.

$$Z \sim N(\mu = 0, \sigma = 1)$$

- Sketch Z curve
- sd determined by POI
- estimate probabilities via empirical rule
- 50/50
- 2.35%, 13.5%, 34%
- $\mathbb{P}(Z < 0) = .0235$
- $\mathbb{P}(Z < -2) = .0235$
- $\mathbb{P}(Z > 1) = .0235 + .135 = 0.1585$

Calculating Probabilities for a Standard Normal Distribution

Three ways to calculate probabilities from a normal distribution:

(1) ~~Calculus~~

(2) Normal probability table

- The textbook has a normal probability table in Appendix B.1, which is included as the next two pages.

(3) R

- $\mathbb{P}(Z \leq q)$ =
`pnorm(q, mean = 0, sd = 1, lower.tail = TRUE)`

Example 3.1. ***Calculating standard normal probabilities practice***
Let Z be a standard normal random variable, $Z \sim N(\mu = 0, \sigma = 1)$.
Calculate the following probabilities. Include sketches of the normal curves with the probability areas shaded in.

(1) $\mathbb{P}(Z < 2.67)$

$$\mathbb{P}(Z < 2.67) = 0.9962$$

`pnorm(2.67) = 0.9962074`

(2) $\mathbb{P}(Z > -0.37)$

$$\mathbb{P}(Z > -0.37) = 1 - \mathbb{P}(Z < -0.37) = 1 - 0.3557 = 0.6443$$

`pnorm(-0.37, lower.tail = F) = 0.6443088`

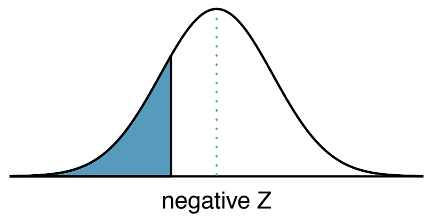
(3) $\mathbb{P}(-2.18 < Z < 2.46)$

$$= \mathbb{P}(Z < 2.46) - \mathbb{P}(Z < -2.18) = 0.9931 - 0.0146 = 0.9785$$

`pnorm(2.46) - pnorm(-2.18) =  0.9784244`

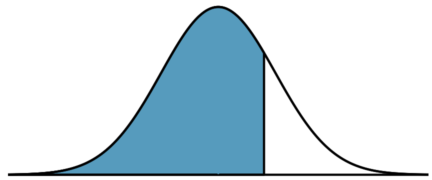
(4) $\mathbb{P}(Z = 1.53)$

$$\mathbb{P}(Z = 1.53) = \mathbb{P}(Z < 1.53) - \mathbb{P}(Z < 1.53) = 0.9370 - 0.9370 = 0$$



Second decimal place of Z										
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Z
0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	−3.4
0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005	−3.3
0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	−3.2
0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010	−3.1
0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013	−3.0
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	−2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	−2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	−2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	−2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	−2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	−2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	−2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	−2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	−2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	−2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	−1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	−1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	−1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	−1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	−1.5
0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808	−1.4
0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968	−1.3
0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151	−1.2
0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357	−1.1
0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587	−1.0
0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841	−0.9
0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119	−0.8
0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420	−0.7
0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743	−0.6
0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085	−0.5
0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446	−0.4
0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821	−0.3
0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207	−0.2
0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602	−0.1
0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000	−0.0

*For $Z \le -3.50$, the probability is less than or equal to 0.0002.



positive Z

	Second decimal place of Z									
Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

*For $Z \geq 3.50$, the probability is greater than or equal to 0.9998.

Calculating Probabilities for a General Normal Distribution

Example 3.2. *Let X be a normal r.v. with mean 8 and standard deviation 2. Calculate $\mathbb{P}(x > 10)$.*

Blue areas (probabilities) do not change via transformation!

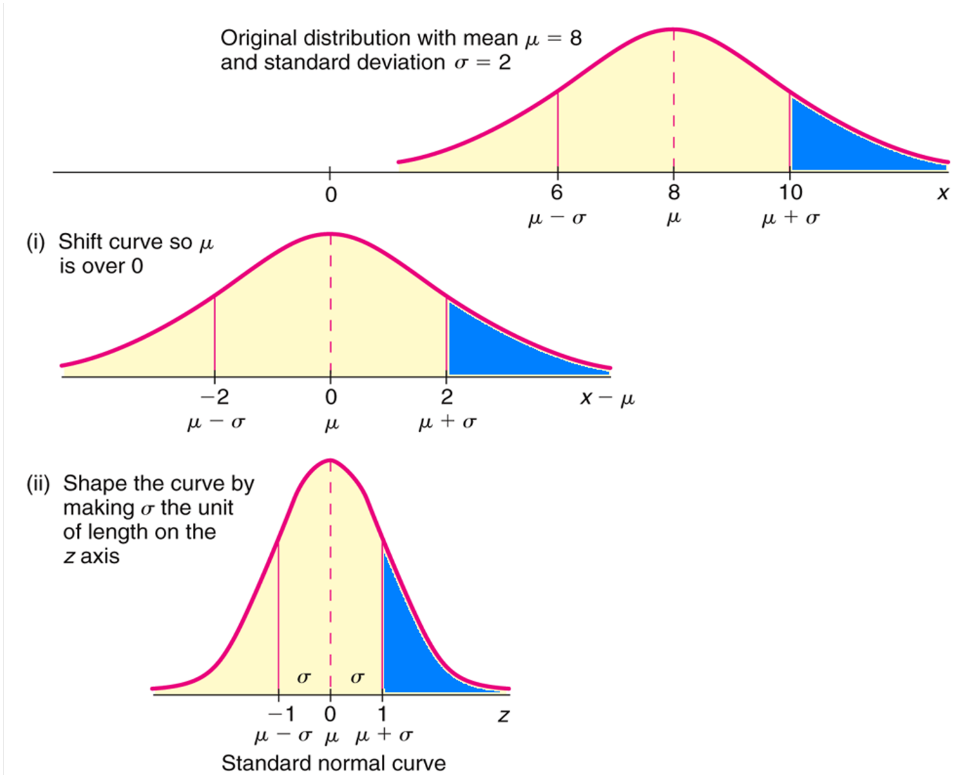


FIGURE 4. Transformation from general normal X to standard normal Z

Z-scores

$$z = \frac{x - \mu}{\sigma}$$

- $\mathbb{P}(X > 10) = \mathbb{P}(Z > \frac{10-8}{2}) = \mathbb{P}(Z > 1) = 1 - \mathbb{P}(Z < 1) = 1 - 0.8413 = 0.1587$
- pnorm(1, lower.tail = F) = 0.1586553
- pnorm(10, mean = 8, sd =2, lower.tail = F)

- The Z-score of an observation quantifies how far the observation is from the mean, in units of standard deviation(s).
- For example, if an observation has Z-score $z = 3.4$, then the observation is 3.4 standard deviations above the mean.

Example 3.3. ***DBP***

Suppose the distribution of diastolic blood pressure (DBP) in 35- to 44-year old men is normally distributed with mean 80 mm Hg and variance 144 mm Hg.

(1) *Mild hypertension is when the DBP is between 90 and 99 mm Hg. What proportion of this population has mild hypertension?*

$$\mathbb{P}(90 < X < 99) = \mathbb{P}\left(\frac{90-80}{\sqrt{144}} < Z < \frac{99-80}{12}\right) = \mathbb{P}(0.83 < Z < 1.58)$$
$$= \mathbb{P}(Z < 1.58) - \mathbb{P}(Z < 0.83) = 0.9429 - 0.7967 = 0.1462$$

pnorm(99, m = 80, s = 12) - pnorm(90, m = 80, s = 12) = 0.1456556

(2) *What is the 10^{th} percentile of the DBP distribution?*

$$x = \mu + z\sigma = 80 + 12z$$

Look up 0.10 in body of table. Get $z = -1.28$:

$$\mathbb{P}(Z < -1.29) = 0.0985\ (diff = 0.0015)$$
$$\mathbb{P}(Z < -1.28) = 0.1003\ (diff = 0.0003)$$
$$x = \mu + z\sigma = 80 + 12(-1.28) = 64.64\ mmHg$$

```
qnorm(p, mean = 0, sd = 1, lower.tail = TRUE)
```

qnorm(.10, m = 80, s = 12) = 64.62138

(3) *What is the 95^{th} percentile of the DBP distribution?*

$$x = \mu + z\sigma = 80 + 12z$$

Look up 0.95 in body of table. Get $z = 1.645$ (average):

$$\mathbb{P}(Z < 1.64) = 0.9495\ (diff = 0.0005)$$
$$\mathbb{P}(Z < 1.65) = 0.9505\ (diff = 0.0005)$$
$$x = \mu + z\sigma = 80 + 12(1.645) = 99.74\ mmHg$$

qnorm(.95, m = 80, s = 12) = 99.73824
= qnorm(.05, m = 80, s = 12, lower.tail = F)

Normal approximation of the binomial distribution

- Recall that a binomial random variable X counts the total number of successes in n independent trials, each with probability p of a success.
- Probability function for $x = 0, 1, ..., n$:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

- Tedious to compute for large number of trails (n), although doable with software like R.
- As n gets big though, the distribution shape of a binomial r.v. gets more and more symmetric, and can be *approximated by a normal distribution.*

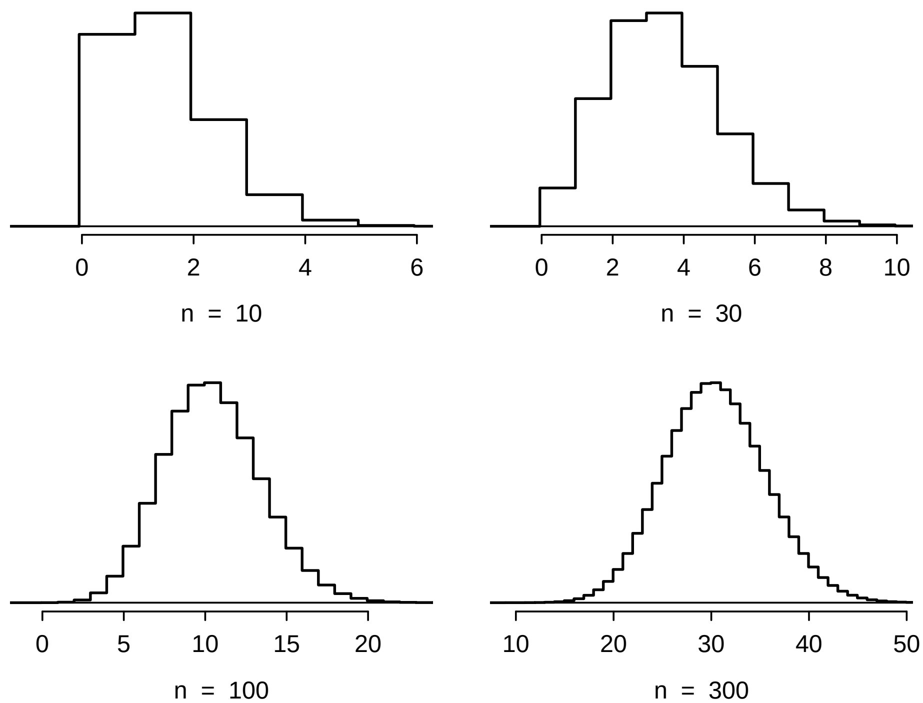
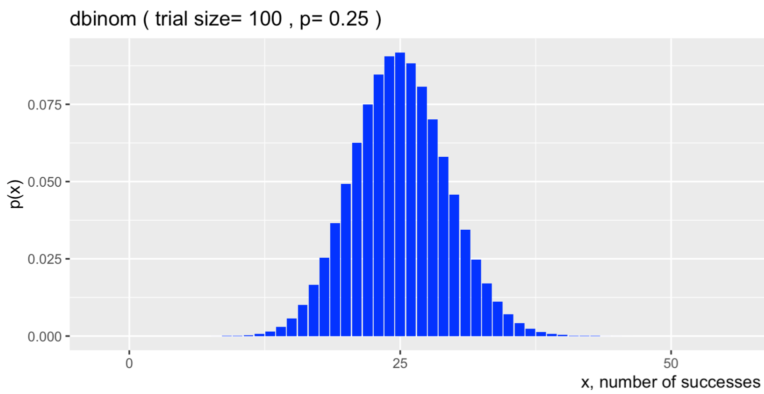


FIGURE 5. Binomial distribution histograms for varying n when $p = 0.10$

Theorem 3.4. ***Normal approximation of the binomial distribution***
The binomial distribution with probability of success p is nearly normal when the sample size n is sufficiently large such that $np \geq 10$ and $n(1-p) \geq 10$.
The approximate normal distribution has parameters corresponding to the mean and standard deviation of the binomial distribution:

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

Example 3.5. ***Vaccinated people testing positive for Covid-19 (revisited)***
About 25% of people that test positive for Covid-19 are vaccinated for Covid-19. Suppose 100 people have tested positive for Covid-19 (independently of each other). Let X denote the number of people that are vaccinated amongst the 100 that tested positive.
What is the probability that fewer than 20 of the people that tested positive are vaccinated?



(1) *Calculate exact probability.*

$$P(X < 20) = P(X \leq 19) = \sum_{k=0}^{19} \binom{100}{k} 0.25^k (0.75)^{100-k}$$

pbinom(19, size=100, prob = .25, lower.tail = TRUE)= 0.09953041

(2) *Calculate approximate probability.*

$$\mu = np = 100(0.25) = 25$$
$$\sigma = \sqrt{npq} = \sqrt{100(0.25)(0.75)} = 4.330127$$

Check conditions: $np = 25 \geq 10, n(1-p) = 75 \geq 10$
CONTINUITY CORRECTION!!!!

$$P(X < 20) = P(X \leq 19.5) = \mathbb{P}\left(Z < \frac{19.5 - 250}{4.33}\right) = \mathbb{P}(Z < -1.27) = 0.1020$$

pnorm(19.5, m = 25, s = 4.330127) = 0.1020119
Without CC:
pnorm(19, m = 25, s = 4.330127) = 0.08292833
pnorm(20, m = 25, s = 4.330127) = 0.1241065

3.3. **Poisson distribution.**

- Discrete distribution
- Used to model count data (# of successes), especially for rare events
- Used to approximate binomial distribution when n is large and p is small

Definition 3.6. *Distribution of a* ***Poisson*** *random variable.*
Let X be the total number of successes in an interval (such as time) with an average success rate λ.
Then probability of observing exactly k successes in a unit interval is

$$P(X = k) = \frac{e^{-\lambda}(\lambda)^k}{k!}, \text{ for } k = 0, 1, 2, \ldots$$

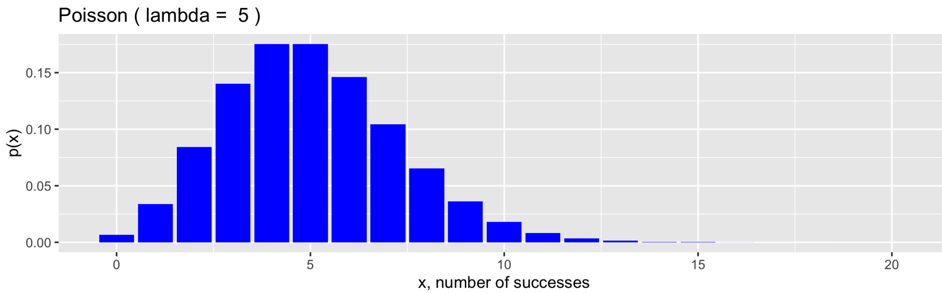
- The parameter of a Poisson distribution is λ.
- If a r.v. X is modeled by a Poisson distribution, then we write in shorthand

$$X \sim \text{Pois}(\lambda)$$

Theorem 3.7. *Mean and SD of a Poisson r.v.*
If X is a Poisson r.v. with parameter λ, then

$$\mu = \mathbb{E}(X) = \lambda$$
$$\sigma^2 = Var(X) = \lambda$$
$$\sigma = SD(X) = \sqrt{\lambda}$$

Example 3.8. ***Typhoid fever***
Suppose there are on average 5 deaths per year from typhoid fever over a 1-year period.



(1) *What is the probability of 3 deaths in a year?*

$$\mathbb{P}(X = 3) = \frac{e^{-5}(5)^3}{3!} = 0.1403739$$

$exp(-L) * (L^k)/factorial(k)$
dpois(x=3, lambda=5)

(2) *What is the probability of 2 deaths in 0.5 years?*

$$\lambda_{t=0.5} = \lambda \cdot t = 5 * 0.5 = 2.5$$
$$\mathbb{P}(X = 2) = \frac{e^{-2.5}(2.5)^2}{2!} = 0.2565156$$

dpois(x=2, lambda=2.5)

(3) *What is the probability of more than 12 deaths in 2 years?*

$$\lambda_{t=2} = \lambda \cdot t = 5 \cdot 2 = 10$$
$$\mathbb{P}(X > 12) = \mathbb{P}(X \geq 13) = \sum_{k=13}^{\infty} \frac{e^{-10}(10)^k}{k!} = 0.2084435$$

$\mathbb{P}(X > 12)$ = *ppois(q=12, lambda=10, lower.tail=F) = 0.2084435*
$1 - \mathbb{P}(X \leq 12)$ = *1 - ppois(q=12, lambda=10, lower.tail=T) = 0.2084435*

Example 3.9. ***Cleft palate***

About 1 in every 1,700 babies is born with cleft palate in the United States. Find the probability that there are 2 babies born with cleft palates amongst the next 3000 births at OHSU.

(1) *Calculate exact probability.*

$$X \sim Binom(n = 3000, p = 1/1700 = 0.0006)$$

$$P(X = 2) = \binom{3000}{2} 0.0006^2 (0.9994)^{2998}$$

dbinom(2, size=3000, prob = 1/1700) = 0.2667187

(2) *Calculate approximate probability.*

Rule of thumb: $\frac{1}{10} \leq npq \leq 10$

$$X \sim Poiss(\lambda = np = 3000 \cdot 1/1700 = 1.764706)$$

Check conditions: $\frac{1}{10} \leq npq = 1.763668 \leq 10$

$$\mathbb{P}(X = 2) = \frac{e^{-1.76}(1.76)^2}{2!} = 0.2664631$$

dpois(x=2, lambda=1.76) = 0.2664631