Special continuous distributions

– Math 161a, Spring 2019, San José State University

Prof. Guangliang Chen

March 21, 2019

Outline

Uniform distribution

Exponential distribution

The Normal distribution

i-Clicker Quiz 6 (extra credit)

Let X be a random variable with pdf

$$f(x) = kx^2, \quad 0 < x < 1$$

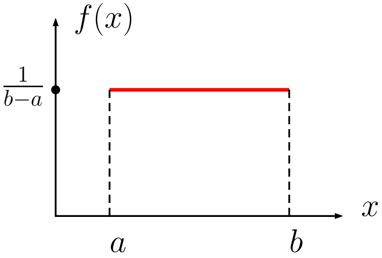
where k is a constant. Find $P(X = \frac{1}{2})$.

- $\frac{1}{8}$
- $\frac{1}{4}$
- $\frac{1}{2}$
- $\frac{3}{4}$
- None of the above

Uniform distribution

Def 0.1 ($X \sim \mathrm{Unif}(a, b)$). We say that a continuous random variable X has a uniform distribution with parameters a, b if it has the following pdf

$$f(x; \underbrace{a, b}_{\text{parameters}}) = \frac{1}{b-a}, \quad a < x < b$$



Ex 0.1. Suppose $X \sim \mathrm{Unif}(2, 4)$. Find the pdf and cdf of X.

Ex 0.2. Suppose a bus arrives at a stop uniformly random between noon and 12:15pm, and you arrive at the bus stop exactly at noon. What is the probability that you will wait

(1) no more than 5 minutes or

(2) between 5 and 10 minutes, or

(3) more than 10 minutes?

Theorem 0.1. *If* $X \sim \mathrm{Unif}(a, b)$*, then the cdf is*

$$F(x) = \frac{x-a}{b-a}, \quad a < x < b.$$

and the mean and variance are

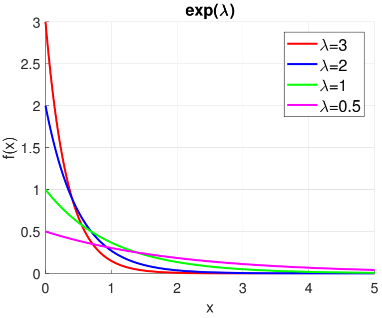
$$E(X) = \frac{a+b}{2}, \quad Var(X) = \frac{(b-a)^2}{12}$$

Exponential distribution

Def 0.2 ($X \sim \text{Exp}(\lambda)$)**.** We say that a continuous random variable X has an exponential distribution with parameter λ if its pdf is given by

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x > 0$$

(Verify that $\int_0^\infty \lambda e^{-\lambda x}\,\mathrm{d}x = 1$)



Ex 0.3. Common examples are **waiting time** for an event, and **life time** of an electronic device such as light bulb and car battery.

Theorem 0.2. *If* $X \sim \mathrm{Exp}(\lambda)$, *then*

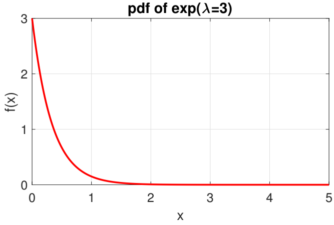
$$E(X) = \frac{1}{\lambda}, \quad Var(X) = \frac{1}{\lambda^2}.$$

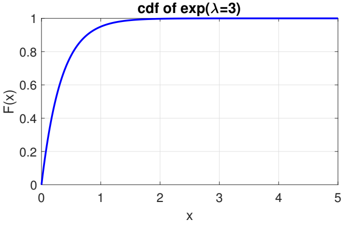
Remark. This shows that $\frac{1}{\lambda}$ is average waiting/life time (λ is the rate).

Ex 0.4. Suppose the life time of a certain brand of light bulbs is exponentially distributed with an average of $1,000$ hours. What is the probability that a new light bulb can exceed this amount of time? Can last between $1,000$ and $2,000$ hours?

Proposition 0.3. *Let $X \sim \mathrm{Exp}(\lambda)$. Then the cdf of X is*

$$F(x;\lambda) = 1 - e^{-\lambda x}, \quad x > 0$$





Complementary cdf function

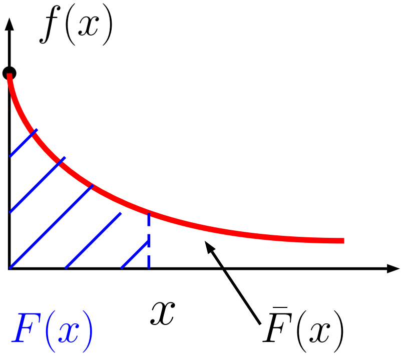
Def 0.3. The complementary cdf of a random variable X is defined as

$$\bar{F}(x) = P(X > x) = 1 - F(x)$$

Remark. If $X \sim \mathrm{Exp}(\lambda)$, then

$$\bar{F}(x) = e^{-\lambda x}, \quad x > 0.$$

It can be thought of as the probability of lasting longer than x hours for a light bulb.



Theorem 0.4 (The memoryless property). *If* $X \sim \mathrm{Exp}(\lambda)$, *then*

$$P(X > t_0 + t \mid X > t_0) = P(X > t), \quad \textit{for any } t_0, t > 0.$$

Interpretation (in the setting of light bulbs):

- $P(X > t)$: probability that a new light bulb can exceed t hours
- $P(X > t_0 + t \mid X > t_0)$: probability that a light bulb can last for t more hours given that it has worked for t_0 hours.

Remark. The exponential distribution is the only continuous distribution that has the memoryless property.

Ex 0.5. Jones figures that the total number of thousands of miles that an auto can be driven before it would need to be junked is an exponential random variable with parameter $\lambda = 1/20$. Smith has a used car that he claims has been driven only 10,000 miles. If Jones purchases the car, what is the probability that she would get at least 20,000 additional miles out of it?

Ex 0.6 (Cont'd)**.** Repeat under the assumption that the lifetime mileage of the car is not exponentially distributed but rather is (in thousands of miles) uniformly distributed over (0, 40).

i-Clicker Quiz 7 (extra credit)

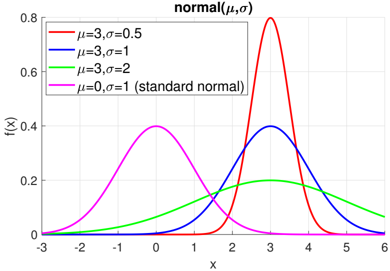
Which of the following functions can be directly used as a pdf (i.e., without any modification)?

- $f_1(x) = e^{-x}, x > 0$
- $f_2(x) = e^{-2x}, x > 0$
- $f_3(x) = e^{-3x}, x > 0$
- All of the above
- None of the above

The Normal distribution

Def 0.4 ($X \sim N(\mu, \sigma^2)$)**.** We say that a continuous random variable X has a normal distribution with parameter μ, σ if it has the following pdf:

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty$$



(1) The normal density curves are **symmetric** (thus also median $= \mu$), **unimodal**, and **bell-shaped**;

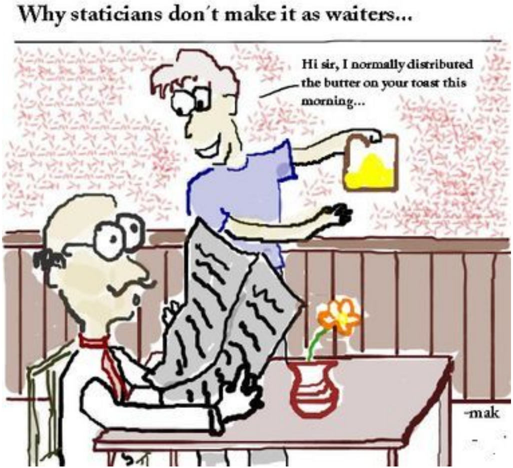
(2) $\mathrm{E}(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$;

(3) We call $N(0, 1)$ standard normal.

$f(x; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R}$

Normal is the most important distribution in probability and statistics:

- Empirically, measurements of a large population often have normal distributions, such as
  - repeated measurements of the same object,
  - heights/weights of a large population, and
  - exam scores of a large class.
- Theoretically, one can show that averages/sums of many independent random variables (individually not necessarily normally distributions) have approximate normal distributions (Central Limit Theorem)


Why staticians don't make it as waiters...
Hi sir, I normally distributed the butter on your toast this morning...
-mak

Bad news: cdfs of normal distributions do not have explicit formulas: For any given point x_0,

$$F(x_0; \mu, \sigma) = P(X \leq x_0) = \int_{-\infty}^{x_0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, \mathrm{d}x$$

Through a change of variable

$$z = \frac{x - \mu}{\sigma} \quad \left(\text{and change of limit } z_0 = \frac{x_0 - \mu}{\sigma}\right)$$

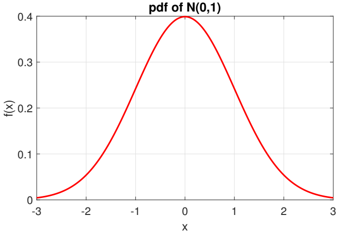
we can obtain that

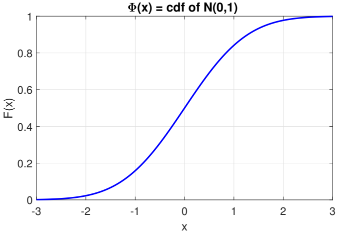
$$F(x_0; \mu, \sigma) = \int_{-\infty}^{z_0} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z = F(z_0; 0, 1).$$

Good news: All cdf calculations for normal distributions can be reduced to similar calculations for the standard normal.

Displaying the cdf of $N(0,1)$

The cdf of standard normal $\Phi(x) \equiv F(x;0,1)$ can be numerically calculated through a computer:





and displayed in a (huge) table, called standard normal table (next page).

Table entry for z is the area under the standard Normal curve to the left of z.

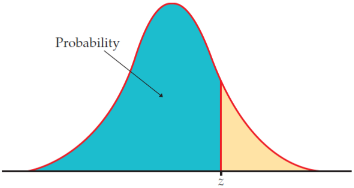


TABLE A

Standard Normal probabilities (continued)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015

Ex 0.7. Suppose $Z \sim N(0,1)$. Find

- $P(Z < 0) =$
- $P(Z < -1.3) =$
- $P(Z > 1.3) =$
- $P(-2.5 < Z < 1.5) =$
- $P(-1 < Z < 1) =$
- $P(-2 < Z < 2) =$
- $P(-3 < Z < 3) =$

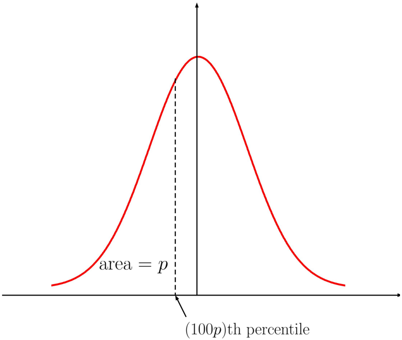
Percentiles

Def 0.5. For any $0 < p < 1$, we define the $(100p)$th percentile of $Z \sim N(0,1)$ as the cutoff z such that

$$p = P(Z < z) = \Phi(z).$$

Alternatively, we may write

$$z = \Phi^{-1}(p).$$



Ex 0.8. Find the 25th (first quartile), 50th (median), 75th (third quartile) percentiles of $Z \sim N(0,1)$.

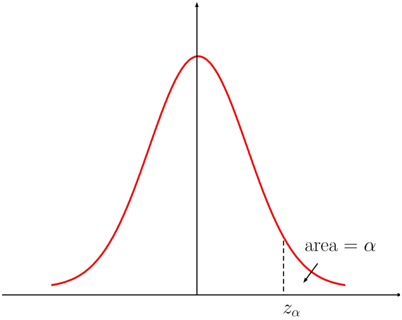
Critical values

Def 0.6. For $0 < \alpha < 1$, define the z_α critical value as

$$P(Z > z_\alpha) = \alpha.$$

Remark. $z_\alpha = \Phi^{-1}(1-\alpha)$, which is the $100(1-\alpha)\text{th percentile:}$

$$P(Z < z_\alpha) = 1 - \alpha.$$



Ex 0.9. Find z_α for $\alpha = .01, .05, .1$

Standardization

Def 0.7. Let $X \sim N(\mu, \sigma^2)$. The standardized variable is

$$Z = \frac{X - \mu}{\sigma}.$$

Proposition 0.5. *If $X \sim N(\mu, \sigma^2)$, then*

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

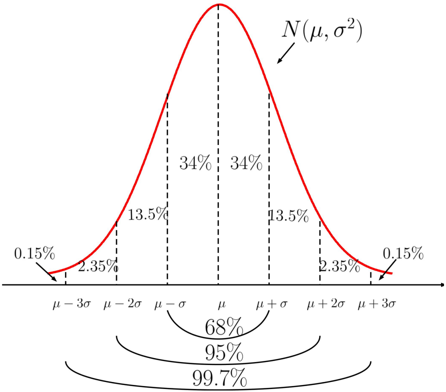
Correspondingly,

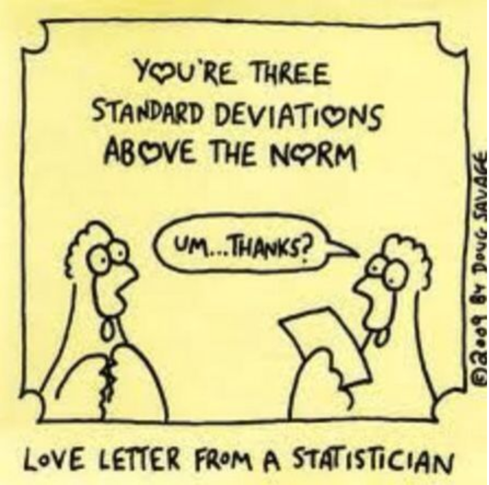
$$F_X(x; \mu, \sigma) = P(X \leq x) = P\left(Z \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right).$$

Ex 0.10. Suppose $X \sim N(5, 3^2)$. Verify that $P(X > 4.1) = 0.6179$, $P(X < -1) = 0.0228$ and $P(2 < X < 5.3) = 0.3812$.

Ex 0.11. Suppose $X \sim N(5, 3^2)$. Find the 90th percentile.

The 68-95-99.7 rule

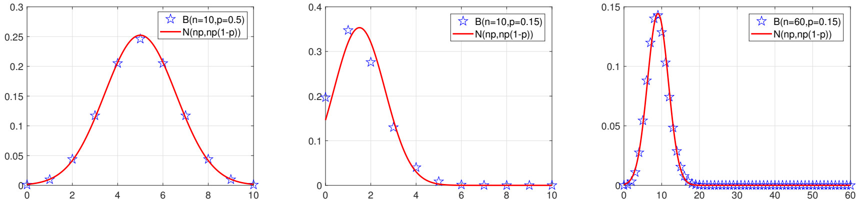



YOU'RE THREE
STANDARD DEVIATIONS
ABOVE THE NORM
UM...THANKS?
©2009 BY DOUG SAVAGE
LOVE LETTER FROM A STATISTICIAN

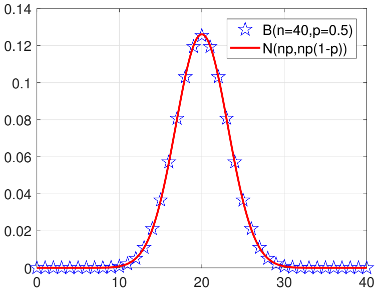
Normal approximation to binomial

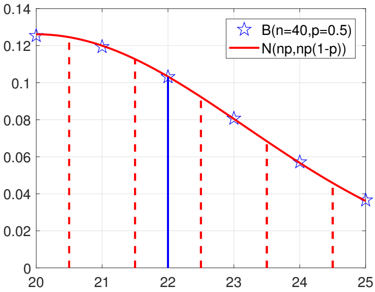
Theorem 0.6. *Let $X \sim B(n,p)$. Then for large n (i.e., $np, n(1-p) \geq 10$),*

$$X \overset{\text{approx}}{\sim} N(np, np(1-p)), \quad \textit{or equivalently, } \frac{X - np}{\sqrt{np(1-p)}} \overset{\text{approx}}{\sim} N(0,1).$$



Ex 0.12. Use normal approximation to find the probability of getting exactly 22 heads when tossing a fair coin 40 times (answer: Binomial 0.1031, Normal 0.1030). What about no more than 22 heads (answer: Binomial 0.7852, Normal approximation 0.7364, and Normal+continuity correction 0.7854)?


B(n=40,p=0.5)
N(np,np(1-p))
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0
10
20
30
40


B(n=40,p=0.5)
N(np,np(1-p))
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
20
21
22
23
24
25