Lec 18

discrete: $\sum x\, p(x)$

4.2.2 Expected Values

For a discrete random variable X, $E(X)$ was obtained by summing $x \cdot p(x)$ over possible X values. Here we replace summation by integration and the pmf by the pdf to get a continuous weighted average.

> **Definition 20.** The expected or mean value of a continuous random variable X with pdf $f(x)$ is $E[X] = \int_{-\infty}^{\infty} x f(x)\, dx$

NOTE: $E(X)$ is the most frequently used measure of population location or center.

Example 58. (Example 57 continued) The pdf of weekly gravel sales X was

$$f(x) = \begin{cases} \frac{3}{2}(1 - x^2) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

So

$$E(X) = \int_0^1 \underbrace{x \cdot \frac{3}{2}(1 - x^2)}_{\frac{3}{2}x - \frac{3}{2}x^3}\, dx = \left(\frac{3}{4}x^2 - \frac{3}{8}x^4\right)\Big|_0^1 = \frac{3}{4} - \frac{3}{8} = \frac{3}{8}$$

Often we wish to compute the expected value of some function $h(X)$ of the random variable X.

> If X is a continuous random variable with pdf $f(x)$ and $h(X)$ is any function of X, then
>
> $$E[h(X)] = \int_{-\infty}^{\infty} h(x) f(x)\, dx$$

> **Definition 21.** The **variance** of a continuous random variable X with pdf $f(x)$ and mean value μ is
>
> $$\sigma_X^2 = \mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$$
>
> The **standard deviation** (SD) of X is
>
> $$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

The variance and standard deviation give quantitative measures of how much spread there is in the distribution or population of x values.

> Shortcut formula
>
> $Var(X) = E[X^2] - E[X]^2$

Example 59. (Example 57 continued) For X = weekly gravel sales, we computed $E(X) = \frac{3}{8}$. Since

$$E(X^2) = \int_0^1 x^2 \cdot \frac{3}{2}(1-x^2)\,dx$$

$$= \int_0^1 \left(\frac{3}{2}x^2 - \frac{3}{2}x^4\right)dx$$

$$= \left(\frac{1}{2}x^3 - \frac{3}{10}x^5\right)\Big|_0^1 = \frac{1}{2} - \frac{3}{10} = \frac{2}{10} = \frac{1}{5}$$

$f(x) = \frac{3}{2}(1-x^2)$

$0 \leq x \leq 1$

Then

$$\text{Var}(X) = \frac{1}{5} - \left(\frac{3}{8}\right)^2$$

and

$$\sigma_X = \sqrt{\frac{1}{5} - \left(\frac{3}{8}\right)^2}$$

> When $h(X) = aX + b$, the expected value and variance of $h(X)$ satisfy the same properties as in the discrete case:
>
> $$E[h(X)] = aE[X] + b$$
>
> and
>
> $$\text{Var}[h(X)] = a^2 Var(X)$$
>
> $\sigma_{h(X)} = |a|\sigma_X$

4.3 The Normal Distribution / Gaussian Distribution

The normal distribution is the most important one in all of probability and statistics.

> **Definition 22.** A continuous random variable X is said to have a **normal distribution** with parameters μ and σ^2, where $-\infty < \mu < \infty$ and $\sigma > 0$, if the pdf of X is
>
> $$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$
>
> The statement that X is normally distributed with parameters μ and σ^2 is often abbreviated $N(\mu, \sigma^2)$

It can be shown that for $X \sim N(\mu, \sigma^2)$

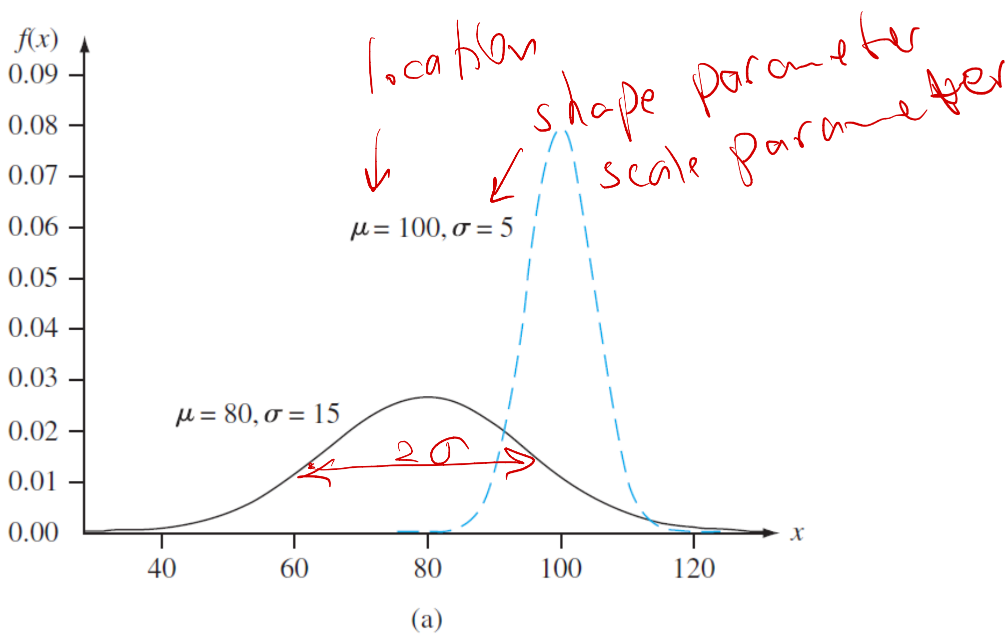
$$E(X) = \mu$$

and

$$\mathrm{Var}(X) = \sigma^2$$

Figure below presents graphs of $f(x; \mu, \sigma^2)$ for several different (μ, σ^2) pairs.

- Each density curve is symmetric about μ and bell-shaped, so the center of the bell (point of symmetry) is both the mean of the distribution and the median.
- The mean μ is a location parameter, since changing its value shifts the density curve.
- σ^2 is referred to as a scale parameter, because changing its value stretches or compresses the curve.
- The inflection points of a normal curve (points at which the curve changes from turning downward to turning upward) occur at $\mu - \sigma$ and $\mu + \sigma$.



(a)

Two different normal density curves

4.3.1 The Standard Normal Distribution

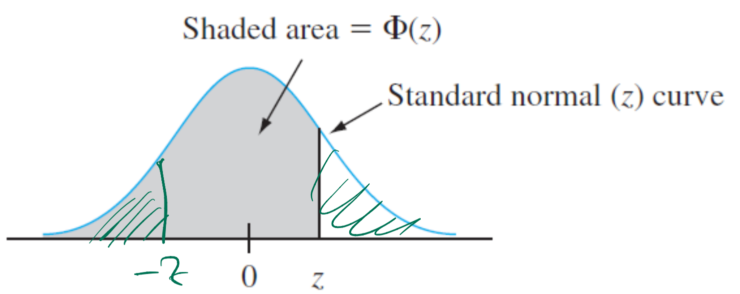
Definition 23. The normal distribution with parameter values $\mu = 0$ and $\sigma = 1$ is called the **standard normal distribution**. A random variable having a standard normal distribution is called a **standard normal random variable** and will be denoted by Z. The pdf of Z is

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \qquad Z \sim N(0,1)$$

error function

CDF: $\Phi(z) = P(Z \le z)$

Appendix Table A.3 gives ______________, the area under the standard normal density curve to the left of z for selected z's. Figure below illustrates the type of cumulative area (probability) tabulated in Table A.3.



Standard normal cumulative areas tabulated in Appendix Table A.3

Example 60. Let's determine the following standard normal probabilities:

(a) $P(Z \le 1.25)$; ≈ 0.8944

(b) $P(Z > 1.25)$; $= P(Z \ge 1.25) = 1 - 0.8944$

(c) $P(Z \le -1.25)$; $= P(Z \ge 1.25) = 1 - 0.8944$

(d) $P(-0.38 \le Z \le 1.25)$; $= P(Z \le 1.25) - \underbrace{P(Z \le -0.38)}$

(e) $P(Z \le 5)$ ≈ 1

Solution.

$P(X \le -5) \approx 0$

$P(Z \ge 0.38)$

$= 1 - P(Z \le 0.38)$

$= 0.8944 - (1 - 0.648)$

$\phi(1.12) = P(Z \le 1.12) \approx 0.8686$

Table A.3 Standard Normal Curve Areas *(cont.)* $\Phi(z) = P(Z \le z)$

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
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9278	.9292	.9306	.9319

4.3.2 Percentiles of the Standard Normal Distribution

For any p between 0 and 1, Appendix Table A.3 can be used to obtain the $(100p)$th percentile of the standard normal distribution.

In general, the $(100p)$th percentile is identified by the row and column of Appendix Table A.3 in which the entry p is found (e.g., the 67th percentile is obtained by finding .6700 in the body of the table, which gives $z = 0.44$).

If p does not appear, the number closest to it is typically used, although linear interpolation gives a more accurate answer. For example, to find the 95th percentile, look for .9500 inside the table. Although it does not appear, both .9495 and .9505 do, corresponding to $z = 1.64$ and 1.65, respectively. Since .9500 is halfway between the two probabilities that do appear, we will use 1.645 as the 95th percentile.

Example 61. Find the 99th percentile of the standard normal distribution.

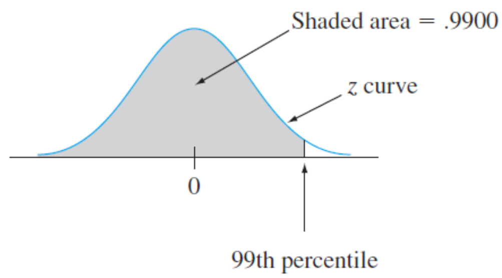
Solution.

$P(Z \le z) = 0{,}01$

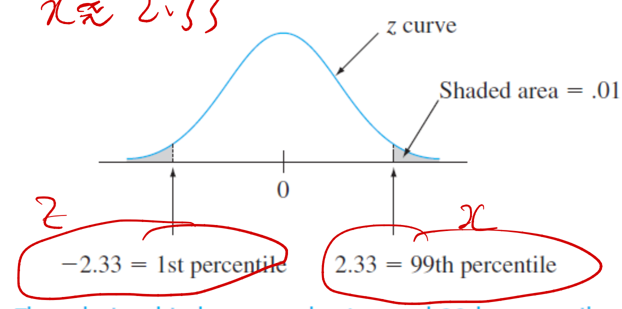
?

$P(Z \ge x) = 0{,}01$ $\quad z = -x = -2.33$

$x \approx 2{,}33$



Finding the 99th percentile



The relationship between the 1st and 99th percentiles

Table A.3 Standard Normal Curve Areas *(cont.)* $\Phi(z) = P(Z \le z)$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936

$P(Z \le 2.33) = 0.9901$