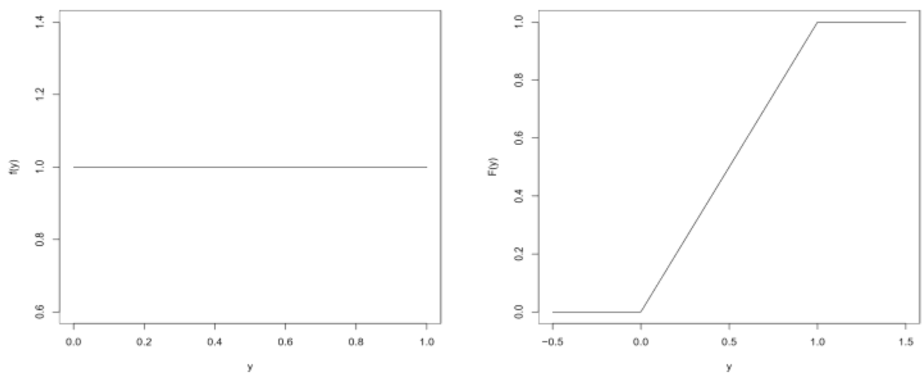
Section 4.6 Normal distribution

Thursday, October 20, 2016 9:41 AM

Section 4.6
Normal di...



pdf, $f_Y(y)$ cdf, $F_Y(y)$

Figure 4.9: *The* $\mathcal{U}(0,1)$ *probability density function and cumulative distribution function.*

4.6 Normal distribution

TERMINOLOGY: A random variable Y is said to have a **normal distribution** if its pdf is given by

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, & -\infty < y < \infty \\ \cancel{0,} & \cancel{\text{otherwise}} \end{cases}$$

$f_Y(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$

$-\infty < y < \mu$

Shorthand notation is $Y \sim \mathcal{N}(\mu, \sigma^2)$. There are two parameters in the normal distribution: the mean $E(Y) = \mu$ and the variance $V(Y) = \sigma^2$.

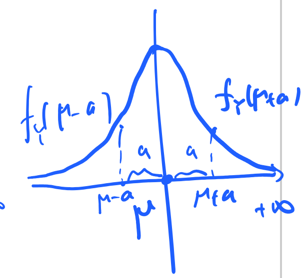
$\left(\frac{y-\mu}{\sigma}\right)^2$ $\quad \frac{(\mu-a-\mu)^2}{\sigma^2} = \frac{a^2}{\sigma^2}$

FACTS:

(a) The $\mathcal{N}(\mu, \sigma^2)$ pdf is symmetric about μ; that is, for any $a \in \mathcal{R}$,

$$f_Y(\mu - a) = f_Y(\mu + a).$$

$\frac{(\mu+a-\mu)^2}{\sigma^2} = \frac{a^2}{\sigma^2}$ $\quad -\infty$



(b) The $\mathcal{N}(\mu, \sigma^2)$ pdf has points of inflection located at $y = \mu \pm \sigma$ (verify!).

(c) $\lim_{y\to\pm\infty} f_Y(y) = 0$.

TERMINOLOGY: A normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called the **standard normal distribution**. It is conventional to let Z denote a random variable that follows a standard normal distribution; we write $Z \sim \mathcal{N}(0,1)$.

$Y \sim N(\mu, \sigma^2)$

IMPORTANT: Tabled values of the standard normal probabilities are given in Appendix III (Table 4, pp 848) of WMS. This table turns out to be helpful since the integral

$$F_Y(y) = P(Y \le y) = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2} dt$$

does not exist in closed form. Specifically, the table provides values of

$$1 - F_Z(z) = P(Z > z) = \int_z^\infty f_Z(u)du,$$

$\mu = 0$
$\sigma = 1$

where $f_Z(u)$ denotes the nonzero part of the standard normal pdf; i.e.,

$$f_Z(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}.$$

$\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right\}$

To use the table, we need to first prove that any $\mathcal{N}(\mu, \sigma^2)$ distribution can be "transformed" to the (standard) $\mathcal{N}(0,1)$ distribution (we'll see how to do this later). Once we do this, we will see that there is only a need for one table of probabilities. Of course, probabilities like $F_Y(y) = P(Y \le y)$ can be obtained using software too.

$dy = \sigma dz$

Example 4.10. Show that the $\mathcal{N}(\mu, \sigma^2)$ pdf integrates to 1.

$z = \frac{y-\mu}{\sigma}$ $dz = \frac{1}{\sigma} dy$

Proof. Let $z = (y - \mu)/\sigma$ so that $dz = dy/\sigma$ and $dy = \sigma dz$. Define

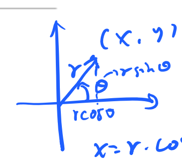
$$\begin{aligned} I &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2} dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}z^2} \sigma dz \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz. \end{aligned}$$

$\int_{-\infty}^{+\infty} e^{-z^2/2} dz = \sqrt{2\pi}$

We want to show that $I = 1$. Since $I > 0$, it suffices to show that $I^2 = 1$. Note that

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-\left(\frac{x^2+y^2}{2}\right)\right] dxdy. = 1 \end{aligned}$$

Switching to polar coordinates; i.e., letting $x = r\cos\theta$ and $y = r\sin\theta$, we get $x^2 + y^2 = r^2(\cos^2\theta + \sin^2\theta) = r^2$, and $dxdy = rdrd\theta$; i.e., the Jacobian of the transformation from



$x = r \cdot \cos\theta$
$y = r \cdot \sin\theta$

(x, y) space to (r, θ) space. Thus, we write

$r = \sqrt{x^2 + y^2}$

$$\begin{aligned}
I^2 &= \frac{1}{2\pi}\int_{\theta=0}^{2\pi}\int_{r=0}^{\infty} e^{-r^2/2} r\,dr\,d\theta \\
&= \frac{1}{2\pi}\int_{\theta=0}^{2\pi}\left[\int_{r=0}^{\infty} re^{-r^2/2}dr\right]d\theta \\
&= \frac{1}{2\pi}\int_{\theta=0}^{2\pi}\left[-e^{-r^2/2}\Big|_{r=0}^{\infty}\right]d\theta \\
&= \frac{1}{2\pi}\int_{\theta=0}^{2\pi} 1\,d\theta = \frac{\theta}{2\pi}\Big|_{\theta=0}^{2\pi} = 1. \quad \square
\end{aligned}$$

$\int_0^\infty r e^{-r^2/2}\,dr$

$\int_0^\infty e^{-\frac{r^2}{2}}\,d\frac{r^2}{2}$

$d\frac{r^2}{2} = \frac{2r}{2}dr = r\,dr$

$\int_0^\infty e^{-t}\,dt \qquad t = \frac{r^2}{2}$

$= -e^{-t}\Big|_0^\infty = 0 - (-1) = 1$

NORMAL MGF: Suppose that $Y \sim \mathcal{N}(\mu, \sigma^2)$. The mgf of Y is

$$m_Y(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right).$$

$Z \sim N(0,1)$

$M_Z(t) = \exp(0 + \frac{t^2}{2}) = \exp(\frac{t^2}{2})$

Proof. Using the definition of the mgf, we have

$f_Y(y)$

$$\begin{aligned}
m_Y(t) = E(e^{tY}) &= \int_{-\infty}^{\infty} e^{ty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}dy \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{\underbrace{ty-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}_{b}}dy.
\end{aligned}$$

Define $b = ty - \frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2$, the exponent in the last integral. We are going to rewrite b in the following way:

$$\begin{aligned}
b &= ty - \frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2 \\
&= ty - \frac{1}{2\sigma^2}(y^2 - 2\mu y + \mu^2) \\
&= -\frac{1}{2\sigma^2}(y^2 - 2\mu y - 2\sigma^2 ty + \mu^2) \\
&= -\frac{1}{2\sigma^2}[\underbrace{y^2 - 2(\mu+\sigma^2 t)y}_{\text{complete the square}} + \mu^2] \\
&= -\frac{1}{2\sigma^2}[y^2 - 2(\mu+\sigma^2 t)y + \underbrace{(\mu+\sigma^2 t)^2 - (\mu+\sigma^2 t)^2}_{\text{add and subtract}} + \mu^2] \\
&= -\frac{1}{2\sigma^2}\{[y - (\mu+\sigma^2 t)]^2\} + \frac{1}{2\sigma^2}[(\mu+\sigma^2 t)^2 - \mu^2] \\
&= -\frac{1}{2\sigma^2}(y-a)^2 + \frac{1}{2\sigma^2}(\cancel{\mu^2} + 2\mu\sigma^2 t + \sigma^4 t^2 - \cancel{\mu^2}) \\
&= -\frac{1}{2\sigma^2}(y-a)^2 + \underbrace{\mu t + \sigma^2 t^2/2}_{=\ c,\ \text{say}},
\end{aligned}$$

$(c-d)^2 = c^2 - 2cd + d^2$

$+d^2 - d^2 + \mu^2$

$c^2 \quad d \quad c$

$= -\frac{1}{2\sigma^2}(y-a)^2 + c$

$a = (\mu + \sigma^2 t)$

$c = \mu t + \frac{\sigma^2 t^2}{2}$

where $a = \mu + \sigma^2 t$. Noting that $c = \mu t + \sigma^2 t^2/2$ is free of y, we have

$$
\begin{aligned}
m_Y(t) &= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^b dy \\
&= \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(y-a)^2+c} dy = \\
&= e^c \left(\underbrace{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(y-a)^2} dy}_{\mathcal{N}(a,\sigma^2)\text{ density}}\right) = e^c,
\end{aligned}
$$

$$e^c \times e^{-\frac{1}{2}\left(\frac{y-a}{\sigma}\right)^2}$$

since the $\mathcal{N}(a, \sigma^2)$ pdf integrates to 1. Now, finally note

$$e^c \equiv \exp(c) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right). \square$$

EXERCISE: Use the mgf to verify that $E(Y) = \mu$ and $V(Y) = \sigma^2$.

$$m_Y(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$$

$$m_Y'(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \times \left(\mu t + \frac{\sigma^2 t^2}{2}\right)'$$

$$= \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \times (\mu + \sigma^2 t)$$

$$m_Y'(0) = \exp(0+0) \times (\mu + 0)$$

$$= 1 \times \mu = \mu$$

$$m_Y''(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)(\mu + \sigma^2 t)^2$$

$$+ \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right) \times \sigma^2$$

$$m_Y''(0) = 1 \times \mu^2 + \sigma^2 = \mu^2 + \sigma^2$$

$$V(Y) = m_Y''(0) - [m_Y'(0)]^2$$

$$= \mu^2 + \sigma^2 - \mu^2 = \sigma^2$$

IMPORTANT: Suppose that $Y \sim \mathcal{N}(\mu, \sigma^2)$. The random variable

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

Proof. Let $Z = (Y - \mu)/\sigma$. The mgf of Z is given by

$$
\begin{aligned}
m_Z(t) = E(e^{tZ}) &= E\left\{\exp\left[t\left(\frac{Y-\mu}{\sigma}\right)\right]\right\} \\
&= E(e^{tY/\sigma - \mu t/\sigma}) \\
&= e^{-\mu t/\sigma} E(e^{tY/\sigma}) \\
&= e^{-\mu t/\sigma} m_Y(t/\sigma) \\
&= e^{-\mu t/\sigma} \exp\left[\mu(t/\sigma) + \frac{\cancel{\sigma^2}(t/\cancel{\sigma})^2}{2}\right] = e^{t^2/2}
\end{aligned}
$$

$$m_Y(t) = E[e^{tY}]$$

which is the mgf of a $\mathcal{N}(0, 1)$ random variable. Thus, by the **uniqueness** of moment generating functions, we know that $Z \sim \mathcal{N}(0, 1)$. $\square$

USEFULNESS: From the last result, we know that if $Y \sim \mathcal{N}(\mu, \sigma^2)$, then the event

$$\{y_1 < Y < y_2\} = \left\{\frac{y_1 - \mu}{\sigma} < \frac{Y - \mu}{\sigma} < \frac{y_2 - \mu}{\sigma}\right\} = \left\{\frac{y_1 - \mu}{\sigma} < Z < \frac{y_2 - \mu}{\sigma}\right\}.$$

As a result,

$$
\begin{aligned}
P(y_1 < Y < y_2) &= P\left(\frac{y_1 - \mu}{\sigma} < Z < \frac{y_2 - \mu}{\sigma}\right) \\
&= F_Z\left(\frac{y_2 - \mu}{\sigma}\right) - F_Z\left(\frac{y_1 - \mu}{\sigma}\right),
\end{aligned}
$$

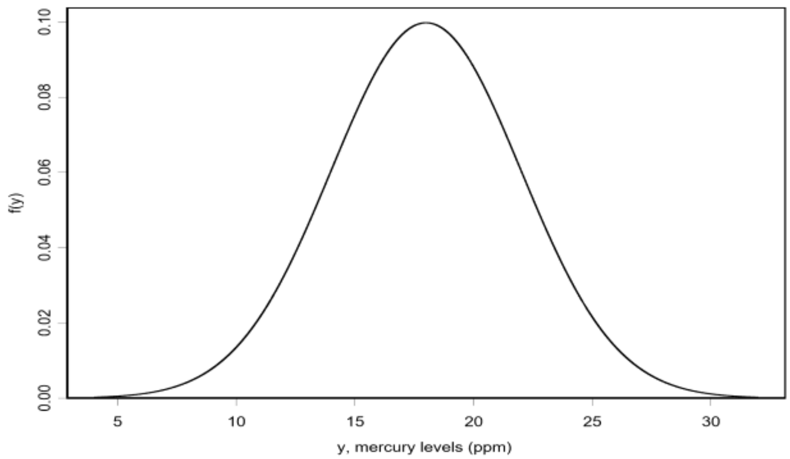


Figure 4.10: *Probability density function, $f_Y(y)$, in Example 4.11. A model for mercury contamination in large-mouth bass.*

Normalcdf

Inv Norm

where $F_Z(\cdot)$ is the cdf of the $\mathcal{N}(0,1)$ distribution. Note also that $F_Z(-z) = 1 - F_Z(z)$, for $z > 0$ (verify!). The standard normal table (Table 4, pp 848) gives values of $1 - F_Z(z)$, for $z > 0$.

Example 4.11. Young large-mouth bass were studied to examine the level of mercury contamination, Y (measured in parts per million), which varies according to a normal distribution with mean $\mu = 18$ and variance $\sigma^2 = 16$, depicted in Figure 4.10.

(a) What proportion of contamination levels are between 11 and 21 parts per million?

SOLUTION. We want $P(11 < Y < 21)$. By standardizing, we see that

$$\begin{aligned} P(11 < Y < 21) &= P\left(\frac{11-18}{4} < \frac{Y-18}{4} < \frac{21-18}{4}\right) \\ &= P\left(\frac{11-18}{4} < Z < \frac{21-18}{4}\right) \\ &= P(-1.75 < Z < 0.75) \\ &= F_Z(0.75) - F_Z(-1.75) = 0.7734 - 0.0401 = 0.7333. \end{aligned}$$

(b) For this model, ninety percent of all contamination levels are above what mercury level?

SOLUTION. We want to find $\phi^Y_{0.10}$, the 10th percentile of $Y \sim \mathcal{N}(18, 16)$; i.e., $\phi^Y_{0.10}$ solves

$$F_Y(\phi^Y_{0.10}) = P(Y \leq \phi^Y_{0.10}) = 0.10.$$

We'll start by finding $\phi^Z_{0.10}$, the 10th percentile of $Z \sim \mathcal{N}(0, 1)$; i.e., $\phi^Z_{0.10}$ solves

$$F_Z(\phi^Z_{0.10}) = P(Z \leq \phi^Z_{0.10}) = 0.10.$$

From the standard normal table (Table 4), we see that

$$\phi^Z_{0.10} \approx -1.28.$$

We are left to solve the equation:

$$\frac{\phi^Y_{0.10} - 18}{4} = \phi^Z_{0.10} \approx -1.28 \Longrightarrow \phi^Y_{0.10} \approx -1.28(4) + 18 = 12.88.$$

Thus, 90 percent of all contamination levels are greater than 12.88 parts per million. □

4.7 The gamma family of distributions

INTRODUCTION: In this section, we examine an important family of probability distributions; namely, those in the gamma family. There are three well-known "named distributions" in this family:

- the exponential distribution
- the gamma distribution
- the χ^2 distribution.

NOTE: The exponential and gamma distributions are popular models for lifetime random variables; i.e., random variables that record "time to event" measurements, such as the lifetimes of an electrical component, death times for human subjects, waiting times in Poisson processes, etc. Other lifetime distributions include the lognormal, Weibull, loggamma, among others.

Important things about Normal

Thursday, October 20, 2016 2:14 PM

Normal $Y \sim N(\mu, \sigma^2)$

① Calculate Probability TI-84 → Distr → normalcdf

$F_Y(y) = P(\underline{\ } Y \le y) = \text{normalcdf}(-10^{99}, y, \mu, \sigma)$

normalcdf (lower bound, upper bound, μ, σ)

not σ^2

$P(a \le Y < b) = \text{normalcdf}(a, b, \mu, \sigma)$

$N(\mu, \sigma^2)$ $\sigma^2 = 4.$ $\sigma = 2$

$\underline{Y \sim N(1, 4)}$ $P(-1 < Y < 3) \approx 0.68$

② $\underline{P(Y \le \phi_p) = p}$ $p = 0.5$

pth percentile $\phi_{0.5}$ median

$\underline{Y \sim N(\mu, \sigma^2)}$ $\phi_{0.5} = \mu$?

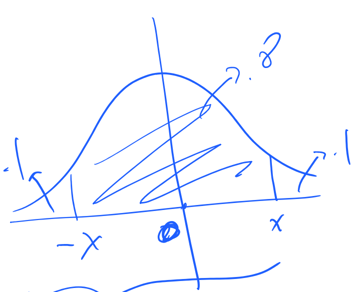
To find ϕ_p for $Y \sim N(\mu, \sigma^2)$ you can use $\underline{\text{invNorm}(p, \mu, \sigma^2)} = \phi_p$

Question. $Z \sim N(0, 1)$

Find the value of x s.t.

$\underline{P(-x < Z < x) = .8}$ how to find x?

by symmetry $P(Z < x) = .9$



So $x = \text{invNorm}(.9, 0, 1)$

Question. $Y \sim N(\mu, 1)$

$\underline{P(Y < 8) = .3}$ what is μ?

$\frac{Y - \mu}{\sqrt{1}} = Y - \mu \sim N(0, 1)$

Idea: $Y \sim N(\mu, 1)$

$P(Y - \mu < 8 - \mu) = P(Z < 8 - \mu)$

$$\frac{Y-\mu}{\sqrt{1}} = Y-\mu \sim N(0,1)$$

Standardize: $Y \sim N(\mu, 1)$

So $0.3 = P(Y<8) = P(\underbrace{Y-\mu}_{N(0,1)} < 8-\mu) = P(Z < 8-\mu)$

To find $8-\mu$ such that $P(Z<8-\mu)$, we can use invNorm

i.e, $8-\mu = \text{invNorm}(.3, 0, 1)$

because $Z \sim N(0,1)$

So $\mu = 8 - \text{invNorm}(.3, 0, 1)$