# ST 501: Fundamentals Statistical Inference

## Continuous Random Variables (part I)

Fall 2024

Department of Statistics, North Carolina State University.

# The uncountable charms of countability

Recall Kolmogorov axioms of probability

- There is a sample space $\Omega$
- There is a collection of subsets $\mathcal{F}$ of $\Omega$, i.e., $\mathcal{F} \subset 2^\Omega$
- There is a non-negative, countably additive function $P \colon \mathcal{F} \mapsto [0, 1]$ with $P(\Omega) = 1$.

A discrete rv is one whose sample space $\Omega$ is a discrete (i.e., countable) subset of the real numbers.

In that case we can let $\mathcal{F} = 2^\Omega$ (the set of all subsets of $\Omega$) and define our probability measure $P$ by assigning non-negative numbers to $P(\omega)$ for all $\omega \in \Omega$, subject to the constraint that

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

This approach, however, fails completely when the sample space for the rv is no longer countable.

The exact details is way outside the scope of this class, but if $A$ is an uncountable set then

$$P(A) = \sum_{\omega \in A} P(\omega)$$

is potentially ill-defined.

If we replace the sum by an integral, i.e.,

$$P(A) = \int_A P(\omega)\,\mathrm{d}\omega$$

then we will need to set $P(\omega) = 0$ for all $\omega \in \Omega$.

Now, assuming there is a $f : \mathbb{R} \mapsto [0, 1]$ such that we can define

$$P(A) = \int_{\omega \in A} f(\omega)\, \mathrm{d}\omega$$

Then, $P(A)$ is well-defined whenever $A$ is a countable union of intervals or a countable intersection of intervals. However there still exists $A \subset \mathbb{R}$ for which the above definition of $P(A)$ is ill-defined. That is to say, we cannot define $P(A)$ for all subsets $A \subset \Omega$.

In summary, we have the following definition of a continuous random variable.

## Definition

A rv $X$ is said to be a continuous rv if $X$ take values in some non-empty interval $S \subset \mathbb{R}$ and satisfy the following conditions

- $P(X \in S) = 1$
- There exists a function $f : S \mapsto [0, 1]$ such that for all intervals $I \subset S$ of the form $I = [a, b], [a, b), (a, b], (a, b)$,

$$P(X \in I) = \int_{x \in I} f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x$$

- For all $A \subset S$, if $A = I_1 \cup I_2 \cup I_3 \cup \dots$ is a countable union of disjoint intervals $I_j \subset S$, then

$$P(A) = P(I_1) + P(I_2) + P(I_3) + \cdots = \sum_{k \geq 1} P(I_k)$$

Important From the previous definition if $X$ is a continuous rv then $P(X = x) = 0$ for all $x \in S$.


6

Important From the previous definition if $X$ is a continuous rv then $P(X = x) = 0$ for all $x \in S$.

Important $P(X = x) = 0$ for all $x$ when $X$ is a continuous rv.


6

Important From the previous definition if $X$ is a continuous rv then $P(X = x) = 0$ for all $x \in S$.

Important $P(X = x) = 0$ for all $x$ when $X$ is a continuous rv.

Important $P(X = x) = 0$ for all $x$ when $X$ is a continuous rv.

Important From the previous definition if $X$ is a continuous rv then $P(X = x) = 0$ for all $x \in S$.

Important $P(X = x) = 0$ for all $x$ when $X$ is a continuous rv.

Important $P(X = x) = 0$ for all $x$ when $X$ is a continuous rv.

And now, for something completely different


6

Important From the previous definition if $X$ is a continuous rv then $P(X = x) = 0$ for all $x \in S$.

Important $P(X = x) = 0$ for all $x$ when $X$ is a continuous rv.

Important $P(X = x) = 0$ for all $x$ when $X$ is a continuous rv.

And now, for something completely different

Important $P(X = x) = 0$ for all $x$ when $X$ is a continuous rv.


6

# pdf and cdf

**Definition**
Let $X$ be a continuous random variable taking values in some non-empty interval $S \subset \mathbb{R}$. The cumulative distribution function (cdf) for $X$ is the function $F\colon S \mapsto [0,1]$ defined by

$$F(x) = P(X \leq x).$$

As $X$ is a continuous r.v., there exists $f\colon \mathbb{R} \mapsto [0,1]$ such that $f(x) = 0$ for all $x \notin S$ and

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(z)\,\mathrm{d}z.$$

$f$ is known as a probability density function (pdf) for $X$.

Note The cdf $F$ is continuous (strictly speaking, differentiable except at some possibly countable number of points) and always non-decreasing, i.e., $F(x) \leq F(z)$ whenever $x \leq z$. Furthermore,

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1$$

$$P(x \leq X \leq z) = F(z) - F(x) \text{ if } x \leq z.$$

Important $f(x)$ is not $P(X = x)$, since if $X$ is a continuous rv then $P(X = x) = 0$. Rather, $f(x)$ is such that, for any $\epsilon > 0$

$$P(x - \epsilon/2 \leq X \leq x + \epsilon/2) \approx \epsilon f(x) \Longrightarrow \lim_{\epsilon \to 0} \frac{P(X - \epsilon/2 \leq X \leq x + \epsilon/2)}{\epsilon}$$

where this "limit" can be taken for (almost) all $x \in S$.

Note. In general, $f$ is not unique. Nevertheless, one can define a canonical via $f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x}$.

Example Let $X$ be a continuous rv with cdf

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

The canonical pdf for $X$ is then

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} = \begin{cases} 0 & x < 0 \\ 1 & 0 \leq x < 1 \\ 0 & 1 \leq x \end{cases}$$

Note Technically speaking, $F(x)$ is not differentiable at $x = 0$ and $x = 1$. Thus the canonical pdf $f(x)$ is only defined on $(-\infty, 0) \cap (0, 1) \cap (1, \infty)$, i.e.,

$$f(x) = \frac{\mathrm{d}F(x)}{\mathrm{d}x} = \begin{cases} 0 & x < 0 \\ 1 & 0 < x < 1 \\ 0 & 1 < x \end{cases}$$

# Expected values for continuous rv

## Definition

Let $X$ be a continuous random variable taking values in $S \subset \mathbb{R}$ with a pdf $f : \mathbb{R} \mapsto [0, 1]$. The expectation of $X$ is defined as

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$$

Furthermore, for any function $g : S \mapsto \mathbb{R}$, the expected value of $g(X)$ is defined as

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) \ \mathrm{d}x$$

Note For ease of exposition we have automatically assume that $f(x) = 0$ whenever $x \notin S$. As such the limits of integration (in the definition of $\mathbb{E}[X]$ and $\mathbb{E}[g(X)]$) are from $-\infty$ to $\infty$.

All the properties of expectation that we saw previously (for discrete random variables), continue to hold for continuous random variables.

All we need to do is to replace summation by integration. For example suppose $X$ is a continuous random variable for which $\mu = \mathbb{E}[X]$ and $\sigma^2 = \mathbb{E}[(X-\mu)^2]$ is well-defined. Then

$$\begin{aligned}
\sigma^2 &= \mathbb{E}[(X-\mu)^2] \\
&= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x - 2\mu \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x + \mu^2 \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \\
&= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.
\end{aligned}$$


12

**99.** A 12-in. bar that is clamped at both ends is to be subjected to an increasing amount of stress until it snaps. Let $Y =$ the distance from the left end at which the break occurs. Suppose $Y$ has pdf

$$f(y) = \begin{cases} \left(\dfrac{1}{24}\right)y\left(1 - \dfrac{y}{12}\right) & 0 \le y \le 12 \\ 0 & \text{otherwise} \end{cases}$$

Compute the following:

**a.** The cdf of $Y$, and graph it.

**b.** $P(Y \le 4)$, $P(Y > 6)$, and $P(4 \le Y \le 6)$

**c.** $E(Y)$, $E(Y^2)$, and $V(Y)$

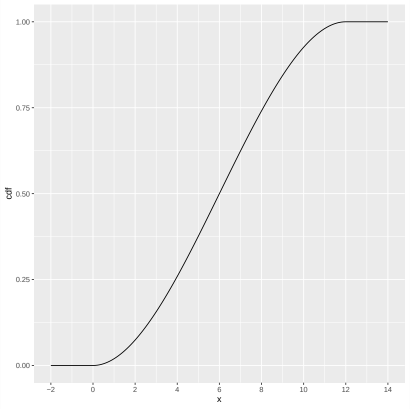
**d.** The probability that the break point occurs more than 2 in. from the expected break point.


13

We have $F(y) = 0$ for $y \leq 0$, $F(y) = 1$ for $y \geq 12$, and

$$F(y) = \frac{1}{24} \int_0^y x\left(1 - \frac{x}{12}\right) \mathrm{d}x = \frac{x^2}{48} - \frac{x^3}{24 \times 36}\Big|_0^y = \frac{y^2}{48} - \frac{y^3}{864}$$

for $0 \leq y \leq 12$.



For part (b), the probabilities are

$$P(Y \leq 4) = F(4) = \frac{16}{48} - \frac{64}{864} = 0.259,$$
$$P(Y > 6) = 1 - F(6) = 1 - \frac{36}{48} + \frac{216}{864} = 0.500,$$
$$P(4 \leq Y \leq 6) = F(6) - F(4) = 0.241$$

For part (c) we have

$$\mathbb{E}[Y] = \frac{1}{24}\int_0^{12} y^2\left(1 - \frac{y}{12}\right)\mathrm{d}y = \frac{y^3}{72} - \frac{y^4}{24 \times 48}\Big|_0^{12} = 6$$
$$\mathbb{E}[Y^2] = \frac{1}{24}\int_0^{12} y^3\left(1 - \frac{y}{12}\right)\mathrm{d}y = \frac{y^4}{96} - \frac{y^5}{24 \times 60} = 43.200,$$
$$\mathrm{Var}[Y] = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = 7.2.$$

Finally, for part (d) we are interested in

$$P(X \leq 4) + P(X \geq 8) = 1 + F(4) - F(8) = 0.519.$$

**Definition**
Let $X$ be a random variable (either discrete or continuous) with cdf $F$. Let $0 < p < 1$. The $p$th percentile of $X$ is defined to be the smallest $x$ such that $F(x) \geq p$, i.e.,

$$F^{-1}(p) = \min\{x\colon P(X \leq x) = F(x) \geq p\}$$

We refer to $F^{-1}\colon (0,1) \mapsto \mathbb{R}$ as the quantile function of $X$. The median of $X$ corresponds to the value of $F^{-1}(0.5)$

# Uniform distribution

**Definition**
Let $a < b$ be finite numbers. A continuous rv $X$ is said to have a uniform distribution on $[a, b]$ if $X$ has cdf

$$F(x) = P(X \leq x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & b < x \end{cases}$$

The canonical pdf for $F$ is then

$$f(x) = \begin{cases} 0 & x < a \\ \frac{1}{b-a} & a \leq x \leq b \\ 0 & b < x \end{cases}$$

We denote a uniform rv on $[a, b]$ by $X \sim \mathrm{Unif}(a, b)$.

**Proposition**

*Let* $X \sim \mathrm{Unif}(a,b)$. *Then*

$$\mathbb{E}[X] = \int_a^b x f(x)\,\mathrm{d}x = \int_a^b \frac{x}{b-a}\,\mathrm{d}x = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2},$$

$$\mathbb{E}[X^2] = \int_a^b \frac{x^2}{b-a}\,\mathrm{d}x = \frac{x^3}{3(b-a)}\Big|_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{1}{3}(b^2+ab+a^2)$$

$$\mathrm{Var}[X] = \frac{1}{3}(a^2+ab+b^2) - \frac{1}{4}(a^2+2ab+b^2) = \frac{1}{12}(b-a)^2.$$


18

# Example: From Poisson to uniform

Let $X_t$ be the random variable for the number of events happening in the time interval $[0, t]$. Suppose we model $X_t$ as a Poisson random variable with rate parameter $\lambda$. Then for any $s < t$, by the infinite divisibility of Poisson rv, $X_t = X_s + X_{t-s}$ where $X_s$ and $X_{t-s}$ are independent Poisson r.v with

$$X_s \sim \mathrm{Pois}(\tfrac{\lambda s}{t}), \qquad X_{t-s} \sim \mathrm{Pois}(\tfrac{\lambda(t-s)}{t})$$

We can interpret $X_s$ and $X_{t-s}$ as the number of events happening in the time interval $[0, s)$ and $[s, t]$, respectively.

Q. Suppose now that $X_t = 1$, i.e., a single event happened during the time $[0, t]$. Let $\mathcal{O}$ be the occurrence time in $[0, t]$ of this event. What is the distribution of $\mathcal{O}$ ?

A. For any $s < t$

$$
\begin{aligned}
P(\mathcal{O} > s \mid X_t = 1) &= P(X_s = 0 \mid X_t = 1) \\
&= \frac{P(X_s = 0, X_t = 1)}{P(X_t = 1)} \\
&= \frac{P(X_s = 0, X_s + X_{t-s} = 1)}{P(X_t = 1)} \\
&= \frac{P(X_s = 0, X_{t-s} = 1)}{P(X_t = 1)} \\
&= \frac{P(X_s = 0) \times P(X_{t-s} = 1)}{P(X_t = 1)} \\
&= \frac{e^{-\lambda s/t} \times e^{-\lambda(t-s)/t} \times \lambda(t-s)/t}{e^{-\lambda t} \times \lambda} = \frac{t-s}{t}.
\end{aligned}
$$


20

We therefore have,

$$P(\mathcal{O} \leq s \mid X_t = 1) = 1 - P(\mathcal{O} > s \mid X_t = 1) = \frac{s}{t}.$$

In other words, given that exactly one event happened during the time interval $[0, t]$, the occurrence time of that event is uniformly distributed on $[0, t]$.


21

# Integral probability transform

The uniform distribution seems too simple to be useful. However, the following result indicates that all continuous rv can be transformed into a uniform rv and vice versa.

**Proposition**
*Let $X$ be a continuous rv with cdf $F$. Then $U = F(X)$ is uniformly distributed on $[0, 1]$. Conversely, $X \sim F^{-1}(\mathrm{Unif}(0, 1))$.*

See also Propositions C and D in section 2.3 of your textbook.


22

Proof Recall that for any $x$, $F(x) = P(X \leq x) \in [0,1]$ and since $F$ is continuous and non-decreasing we can define an inverse function $F^{-1}$ such that for all $t \in [0,1]$,

$$F(F^{-1}(t)) = P_X(X \leq F^{-1}(t)) = t.$$

We therefore have for all $t \in [0,1]$,

$$P_U(U \leq t) = P_X(F(X) \leq t) = P_X(X \leq F^{-1}(t)) = t,$$

i.e., $U \sim \mathrm{Unif}(0,1)$.

Conversely, let $U \sim \mathrm{Unif}(0,1)$ and define $X = F^{-1}(U)$. Then

$$P_X(X \leq x) = P_U(F^{-1}(U) \leq x) = P_U(U \leq F(x)) = F(x).$$


23

**123.** Let $U$ have a uniform distribution on the interval $[0, 1]$. Then observed values having this distribution can be obtained from a computer's random number generator. Let $X = -(1/\lambda)\ln(1 - U)$.

**a.** Show that $X$ has an exponential distribution with parameter $\lambda$. [*Hint*: The cdf of $X$ is $F(x) = P(X \leq x)$; $X \leq x$ is equivalent to $U \leq$ ?]

**b.** How would you use part (a) and a random number generator to obtain observed values from an exponential distribution with parameter $\lambda = 10$?

# Location-scale transform

**Proposition**
*Let $X$ be any continuous rv with pdf $f$. Then for any real numbers $\mu, \sigma$ with $\sigma > 0$, the function*

$$f_{\mu,\sigma}(x) = \frac{1}{\sigma} f((x-\mu)/\sigma)$$

*is also a valid pdf. $f_{\mu,\sigma}$ corresponds to the pdf of $\sigma X + \mu$.*

See also problem 2.62 in your textbook.


25

**Proof** Let $Z = \sigma X + \mu$. Then

$$P(Z \leq z) = P(\sigma X + \mu \leq z) = P(X \leq (z - \mu)/\sigma) = F((z - \mu)/\sigma)$$

Taking derivative of the above with respect to $z$, we obtain

$$f_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z} F((z - \mu)/\sigma) = \frac{1}{\sigma} f((z - \mu)/\sigma).$$

**Example** Let $U \sim \mathrm{Unif}(0, 1)$. For $a < b$, $(b - a)U + a \sim \mathrm{Unif}(a, b)$.



26

# Monotone transformation

**Proposition**

*Let $X$ be a continuous rv with pdf $f$ and $g$ a differentiable monotone function. Then $Z = g(X)$ is a continuous rv with pdf*

$$f_Z(z) = \frac{f(g^{-1}(z))}{|g'(g^{-1}(z))|}$$

See also Proposition B in section 2.3 of your textbok.


27

Proof Suppose that $g$ is non-decreasing. Then

$$P(Z \leq z) = P(g(X) \leq z) = P(X \leq g^{-1}(z)) = F(g^{-1}(z)).$$

Taking derivative of the above expression, we have

$$f_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z} F(g^{-1}(z)) = f(g^{-1}(z)) \times \frac{1}{g'(g^{-1}(z))} = \frac{f(g^{-1}(z))}{|g'(g^{-1}(z))|}$$

since $g'$ is non-negative.


28

For the case where $g$ is non-increasing, we have

$$P(Z \leq z) = P(g(X) \leq z) = P(X \geq g^{-1}(z)) = 1 - F(g^{-1}(z)).$$

The derivative of the above expression is then

$$f_Z(z) = \frac{\mathrm{d}}{\mathrm{d}z}(1 - F(g^{-1}(z))) = -f(g^{-1}(z)) \times \frac{1}{g'(g^{-1}(z))} = \frac{f(g^{-1}(z))}{|g'(g^{-1}(z))|}$$

since $g'$ is now non-positive.


29

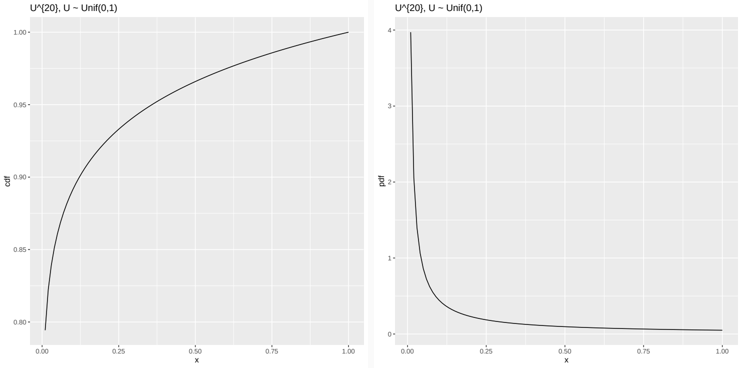
Example Let $U \sim \mathrm{Unif}(0,1)$. Let $X = U^\alpha$ where $\alpha > 0$. Then

$$P(X \leq x) = P(U^\alpha \leq x) = P(U \leq x^{1/\alpha}) = x^{1/\alpha} = F(x)^\alpha$$

where $F$ is the cdf of a $\mathrm{Unif}(0,1)$ rv. We then have

$$f_X(x) = \frac{1}{\alpha} x^{1/\alpha - 1} = \frac{1}{g'(g^{-1}(x))}$$

where $g(x) = x^\alpha$ and hence $g^{-1}(x) = x^{1/\alpha}, g'(x) = \alpha x^{\alpha-1}$




The pdf for $X = U^\alpha$ with $\alpha = 20$ and $U \sim \mathrm{Unif}(0, 1)$ is extremely peaked around $x = 0$, i.e. the cdf for $X = U^\alpha$ increases very fast for $x \approx 0$.

# Example: Cauchy rv

**Definition**
Let $U \sim \text{Unif}(-\pi/2, \pi/2)$. Then $X = \tan U$ is a continuous rv with pdf

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad x \in \mathbb{R}$$

$X$ is known as the Cauchy distribution.


32

# Example: Cauchy rv

**Definition**
Let $U \sim \text{Unif}(-\pi/2, \pi/2)$. Then $X = \tan U$ is a continuous rv with pdf

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad x \in \mathbb{R}$$

$X$ is known as the Cauchy distribution.

Indeed, let $g(x) = \tan x$. Then

$$g'(x) = 1 + (\tan(x))^2,$$
$$g'(g^{-1}(x)) = 1 + (\tan(\arctan(x)))^2 = 1 + x^2.$$

Now the pdf of $U$ is $1/\pi$ and so the result follows.

The Cauchy distribution has several quite interesting properties. The first is that $\mathbb{E}[X]$ does not exists. Indeed

$$\begin{aligned}
\mathbb{E}[X] &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{x}{1+x^2}\mathrm{d}x \\
&= \frac{1}{\pi}\int_{-\infty}^{0} \frac{x}{1+x^2}\mathrm{d}x + \frac{1}{\pi}\int_{0}^{\infty} \frac{x}{1+x^2}\mathrm{d}x \\
&= \frac{1}{\pi}\int_{\infty}^{1} \frac{\mathrm{d}u}{u} + \frac{1}{\pi}\int_{1}^{\infty} \frac{\mathrm{d}u}{u} \\
&= \frac{1}{\pi}\ln u\Big|_{\infty}^{1} + \frac{1}{\pi}\ln u\Big|_{1}^{\infty}
\end{aligned}$$

which is of the form $\infty - \infty$ and is thus ill-defined.

The second interesting property is that if $X$ is Cauchy then $1/X$ is also Cauchy.


33

# Mixture distribution (Optional)

**Proposition**

*Let $F_1, F_2, \ldots, F_K$ be $K$ cdf, and let $p_1, p_2, \ldots, p_K$ be $K$ non-negative numbers with $p_1 + p_2 + \cdots + p_K = 1$. Then*

$$F = \sum_k p_k F_k = p_1 F_1 + p_2 F_2 + \cdots + p_K F_K$$

*is a valid cdf. Suppose furthermore that the $\{F_k\}_{k=1}^K$ are cdf for $K$ continuous rvs. Let $\{f_k\}$ be the corresponding pdf. Then*

$$f(x) = \sum_k p_k f_k$$

*is a valid pdf.*


34

Important Let $X_1, X_2, \ldots, X_K$ be continuous rv with pdf $f_1, f_2, \ldots, f_K$. Then $f(x) = \sum_k p_k f_k$ is not the pdf for the rv $p_1 X_1 + p_2 X_2 + \cdots + p_K X_K$.

To interpret $f(x) = \sum_k p_k f_k$, let $X_1, X_2, \ldots, X_K$ be mutually independent rvs with pdf $f_1, f_2, \ldots, f_K$ and let $Y$ be a discrete rv with pmf $P(Y = k) = p_k$ with $Y$ being independent of the $\{X_k\}$.

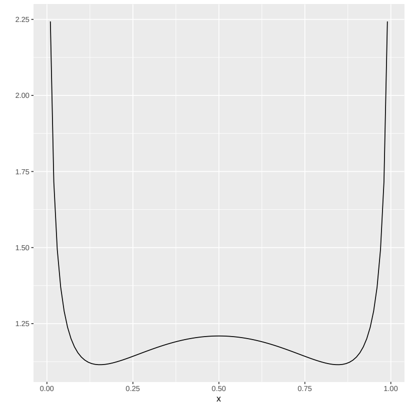
Then $f(x) = \sum_k p_k f_k$ is the pdf for the continuous rv

$$Z = X_1 I(Y = 1) + X_2 I(Y = 2) + \cdots + X_K I(Y = K)$$

where $I(Y = j)$ is the indicator rv for the event $\{Y = j\}$.


35

Example Let $f_1(x) = 1$, $f_2(x) = 6x(1-x)$, $f_3(x) = \frac{1}{(\pi x(1-x))^{1/2}}$.
We visualize $f(x) = 1/3(f_1(x) + f_2(x) + f_3(x))$ below



**113.** The article **"Statistical Behavior Modeling for Driver-Adaptive Precrash Systems" (*IEEE Trans. on Intelligent Transp. Systems*, 2013: 1–9)** proposed the following mixture of two exponential distributions for modeling the behavior of what the authors called "the criticality level of a situation" $X$.

$$f(x; \lambda_1, \lambda_2, p) = \begin{cases} p\lambda_1 e^{-\lambda_1 x} + (1-p)\lambda_2 e^{-\lambda_2 x} & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

This is often called the hyperexponential or mixed exponential distribution. This distribution is also proposed as a model for rainfall amount in **"Modeling Monsoon Affected Rainfall of Pakistan by Point Processes" (*J. of Water Resources Planning and Mgmnt.*, 1992: 671–688).**


37

Example Mixture distributions also allows for mixing a discrete rv and a continuous rv. Suppose we are interested in the amount of coffee consumed by a typical NC State student. If a student does drink coffee then we could model his/her coffee consumption by a uniform random variable over some interval $[0, C]$. However, if a student does not drink coffee then his/her coffee consumption is always 0. If $30\%$ of NC State students do not drink coffee then we have a rv with cdf

$$F(y) = 0.3I(y = 0) + 0.7F_U(y)$$

where $F_U$ is the cdf for a $\mathrm{Unif}([0, C])$ rv.


38

# Normal distribution

> *Everyone believes in the normal law of errors: the mathematicians, because they think it is an experimental fact; and the experimenters, because they suppose it is a theorem of mathematics.*
>
> *— Gabriel Lippman*


39

We now come to the most famous distribution of them all, the Gaussian or normal distribution.

**Definition**
Let $\mu \in \mathbb{R}$ and $\sigma > 0$. A rv $X$ is said to be normally distributed with mean $\mu$ and variance $\sigma^2$ if $X$ has pdf

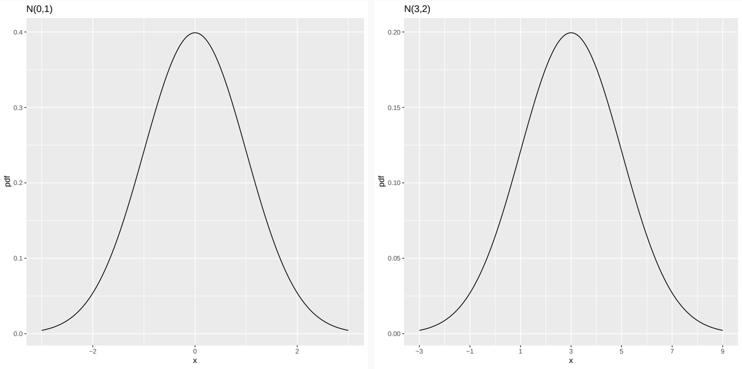
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big), \qquad x \in \mathbb{R}.$$


40

Note We denote a normal rv as $X \sim \mathcal{N}(\mu, \sigma^2)$. If $\mu = 0$ and $\sigma^2 = 1$ then $X \sim \mathcal{N}(0, 1)$ is said to be standard normal. The pdf of a standard normal is

$$f(x) = (2\pi)^{-1/2} e^{-x^2/2}.$$

If $X \sim \mathcal{N}(\mu, \sigma^2)$ then the pdf of $X$ is symmetric around $\mu$, i.e, the median of $X$ is $\mu$.



# Transformation to standard normal

A normal rv is an affine transformation of a standard normal.

**Proposition**
*Let* $X \sim \mathcal{N}(0, 1)$. *Then* $\sigma X + \mu \sim \mathcal{N}(\mu, \sigma^2)$.

The proof follows directly from the form of a pdf under a location-sale transformation. See Proposition A in section 2.3 of your textbook.

The result is very useful as it reduces probability calculation for any normally distributed rv to that for a standard normal.

Example Define $\Phi(x)$ to be the cdf for $N(0,1)$. Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then for any $a < b$

$$P(a \le X \le b) = P\Big(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\Big)$$
$$= \Phi((b-\mu)/\sigma) - \Phi((a-\mu)/\sigma)$$

Important Since the cdf of any normally distributed rv can be reduced to the cdf of a standard normal, it is sufficient to tabulate only the cdf for a standard normal. While $\Phi(x)$ has no close form expression for general $x$, it is nevertheless straightforward to tabulate (see the appendix of your textbook). Several representative values are given below.

$$\Phi(0) = 0.5, \quad \Phi(0.5) \approx 0.691, \quad \Phi(1) \approx 0.84, \quad \Phi(1.5) \approx 0.93,$$
$$\Phi(2) \approx 0.977, \quad \Phi(2.5) \approx 0.994, \quad \Phi(3) \approx 0.9986, \quad \Phi(4) \approx 0.9999$$

Furthermore, since the pdf for $X \sim \mathcal{N}(0, 1)$ is symmetric around 0, i.e., $f(-x) = f(x)$ for all $x \geq 0$, we have, for all $x \geq 0$

$$\Phi(-x) = 1 - \Phi(x), \quad P(|X| \leq x) = P(-x \leq X \leq x) = 2\Phi(x) - 1.$$

Therefore, for $X \sim \mathcal{N}(0, 1)$,

$$P(|X| \leq 1) \approx 0.68, \quad P(|X| \leq 2) \approx 0.95, \quad P(|X| \leq 3) \approx 0.997.$$

Now for $Z \sim \mathcal{N}(\mu, \sigma^2)$, as $Z = \sigma\mathcal{N}(0, 1) + \mu$, we have

$$P(|Z - \mu| \leq \sigma) \approx 0.68,$$
$$P(|Z - \mu| \leq 2\sigma) \approx 0.95,$$
$$P(|Z - \mu| \leq 3\sigma) \approx 0.997.$$

Example The achievement scores for a college entrance exam are normally distributed with mean 75 and standard deviation 10. What fraction of the scores lie between 80 and 90 ?


45

Example The achievement scores for a college entrance exam are normally distributed with mean 75 and standard deviation 10. What fraction of the scores lie between 80 and 90 ?

We say that $Z \sim \mathcal{N}(75, 10)$ and hence, with $X \sim \mathcal{N}(0, 1)$

$$
\begin{aligned}
P(80 \leq Z \leq 90) &= P\Big(\frac{80-75}{10} \leq \frac{Z-75}{10} \leq \frac{90-75}{10}\Big) \\
&= P(0.5 \leq X \leq 1.5) = \Phi(1.5) - \Phi(0.5) \approx 0.242
\end{aligned}
$$


45

Example Suppose the distribution of the height of males is approximately normal with mean 70 inches. Ten percent of individuals are over $6\mathrm{ft} = 72\mathrm{in}$ tall. What is the approximate probability that in a group of 50 people picked at random there will be two or more people who are over $6\mathrm{ft}1\mathrm{in} = 73\mathrm{in}$ tall ?


46

Example Suppose the distribution of the height of males is approximately normal with mean 70 inches. Ten percent of individuals are over $6\mathrm{ft} = 72\mathrm{in}$ tall. What is the approximate probability that in a group of 50 people picked at random there will be two or more people who are over $6\mathrm{ft}1\mathrm{in} = 73\mathrm{in}$ tall ?

Let $Z \sim N(70, \sigma^2)$ be the distribution of the height. We have $P(Z \geq 72) = 0.1$, i.e., $P(Z \leq 72) = 0.9$ and hence

$$P((Z-70)/\sigma \leq 2/\sigma) = 0.9 \implies 2/\sigma = \Phi^{-1}(0.9) \approx 1.28 \implies \sigma \approx 1.56$$

Now let $X$ be the number of people who are over 73 inches tall in a randomly selected group of 50 people. Then $X \sim \mathrm{Bin}(50, p)$ where $p = P(Z \geq 73) \approx 0.027$. We thus have

$$P(X \geq 2) = 1 - P(X \leq 1) \approx 0.3996$$


46

# An ingenious integral identity (optional)

We now verify that the pdf for a $\mathcal{N}(\mu, \sigma^2)$ rv is a valid pdf.

More specifically

$$\begin{aligned}
\Big(\int_{-\infty}^{\infty} e^{-x^2/2}\mathrm{d}x\Big)^2 &= \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x \int_{-\infty}^{\infty} e^{-y^2/2}\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2/2+y^2/2)}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{0}^{\infty}\int_{0}^{2\pi} e^{-r^2/2} r\,\mathrm{d}\theta\,\mathrm{d}r \quad \text{(polar coordinates)} \\
&= 2\pi \int_{0}^{\infty} r e^{-r^2/2}\,\mathrm{d}r = -2\pi e^{-r^2/2}\Big|_0^{\infty} = 2\pi.
\end{aligned}$$

We therefore have

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} \times \sqrt{2\pi} = 1.$$



47

# Mean and variance of normal rv

**Proposition**
*Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then*

$$\mathbb{E}[X] = \mu, \quad \mathrm{Var}[X] = \sigma^2$$


48

# Mean and variance of normal rv

**Proposition**

*Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then*

$$\mathbb{E}[X] = \mu, \quad \mathrm{Var}[X] = \sigma^2$$

Proof We only show that if $X \sim \mathcal{N}(0, 1)$ then $\mathbb{E}[X] = 0$ and $\mathrm{Var}[X] = 1$. The proof that $\mathbb{E}[X] = 0$ follows directly from symmetry around 0 of $X \sim \mathrm{N}(0, 1)$. The proof that $\mathrm{Var}[X] = 1$ follows from integration by parts, i.e.,

$$\begin{aligned}
\mathbb{E}[X^2] &= (2\pi)^{-1/2} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} \, \mathrm{d}x \\
&= -(2\pi)^{-1/2} x e^{-x^2/2} \Big|_{-\infty}^{\infty} + (2\pi)^{-1/2} \int_{-\infty}^{\infty} e^{-x^2/2} \, \mathrm{d}x = 0 + 1 = 1.
\end{aligned}$$

# mgf for normal rvs

**Theorem**
*Let $X$ be a normal rv with mean $\mu$ and variance $\sigma^2$. Then for any $t \in \mathbb{R}$,*

$$\mathbb{E}[e^{tX}] = \exp(\mu t + \tfrac{1}{2}t^2\sigma^2)$$


49

Proof The mgf for $X \sim \mathcal{N}(0,1)$ follows from completing the squares, i.e.,

$$
\begin{aligned}
\mathbb{E}[e^{tX}] &= (2\pi)^{-1/2} \int_{\mathbb{R}} e^{-x^2/2+tx}\,\mathrm{d}x \\
&= (2\pi)^{-1/2} \int_{\mathbb{R}} \exp(-\tfrac{1}{2}(x^2 - 2tx + t^2) + t^2/2)\mathrm{d}x \\
&= e^{t^2/2}(2\pi)^{-1/2} \int_{\mathbb{R}} \exp(-\tfrac{1}{2}(x-t)^2)\mathrm{d}x = e^{t^2/2}.
\end{aligned}
$$

Finally, for $Z \sim \mathcal{N}(\mu, \sigma^2)$, $Z = \sigma X + \mu$ where $X \sim \mathcal{N}(0,1)$ and hence

$$
\mathbb{E}[e^{tZ}] = \mathbb{E}[e^{t\sigma X + t\mu}] = e^{t\mu}\mathbb{E}[e^{(t\sigma)X}] = e^{t\mu} \times e^{(t\sigma)^2/2} = \exp(t\mu + \tfrac{1}{2}t^2\sigma^2).
$$

See Property C in section 4.5 of your textbok

Example Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be independent. What is the distribution of $X = X_1 + X_2$ ?


51

Example Let $X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ be independent. What is the distribution of $X = X_1 + X_2$ ?

We compute the mgf of $X = X_1 + X_2$. We have

$$
\begin{aligned}
m_X(t) &= m_{X_1}(t) \times m_{X_2}(t) \\
&= \exp(t\mu_1 + t^2\sigma_1^2) \times \exp(t\mu_2 + t^2\sigma_2^2) \\
&= \exp(t(\mu_1 + \mu_2) + t^2(\sigma_1^2 + \sigma_2^2)/2)
\end{aligned}
$$

which is the mgf for a $\mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ random variable.

By the uniqueness of mgf, we conclude that $X \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$. See also example G in section 4.5 of your textbook.


51

# Infinite divisiblility of normal rvs (optional)

The previous example can be generalized to yield

**Proposition**
*Let* $X \sim \mathcal{N}(\mu, \sigma^2)$*. Then* $X$ *is infinitely divisible, i.e., for any integer* $n \geq 1$ *we have* $X = X_1 + X_2 + \cdots + X_n$ *where the* $X_i$ *are mutually independent* $\mathcal{N}(\mu/n, \sigma^2/n)$ *random variables.*



52

# Sum of normal rvs

**Proposition**
*Let* $X_1, X_2, \ldots, X_n$ *be mutually independent random variables with* $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$*. Then* $X = X_1 + X_2 + \cdots + X_n$ *is a* $\mathcal{N}(\sum_i \mu_i, \sum_i \sigma_i^2)$ *random variable.*

In simple words, a sum of independent normally distributed random variables is also normally distributed. See also problem 4.89 in your textbook.


53

# Lognormal rv

**Definition**
Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Then $Y = e^X$ is said to have a lognormal distribution with parameters $\mu$ and $\sigma^2$. The pdf for $Y$ is then

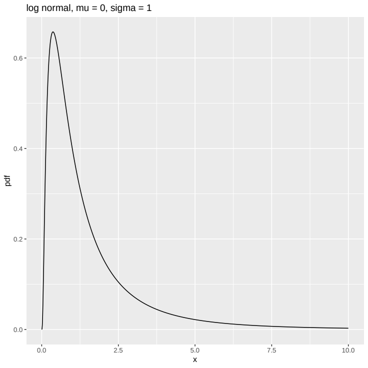
$$f(y) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\log y - \mu)^2}{2\sigma^2}\right)$$

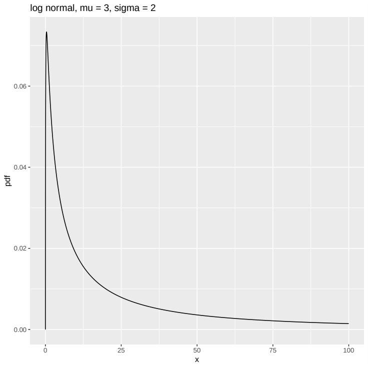
and the cdf for $Y$ can be related to that for $\mathcal{N}(0, 1)$, i.e.,

$$P(Y \leq y) = P((X - \mu)/\sigma \leq (\log y - \mu)/\sigma) = \Phi((\log y - \mu)\sigma)$$

A lognormal rv is skewed towards the right and is usually used to model economic variables such as income or wealth.


54


log normal, mu = 0, sigma = 1
pdf
0.0
0.2
0.4
0.6
0.0
2.5
5.0
7.5
10.0
x


log normal, mu = 3, sigma = 2
pdf
0.00
0.02
0.04
0.06
0
25
50
75
100
x


55

**Proposition**
*Let $Y$ be a log normal rv with parameters $\mu$ and $\sigma^2$. Then*

$$\mathbb{E}[Y] = e^{\mu+\sigma^2/2}, \quad \mathrm{Var}[Y] = e^{2\mu+2\sigma^2} - e^{2\mu+\sigma^2}.$$

Proof Follows directly from the mgf of a normal rv.

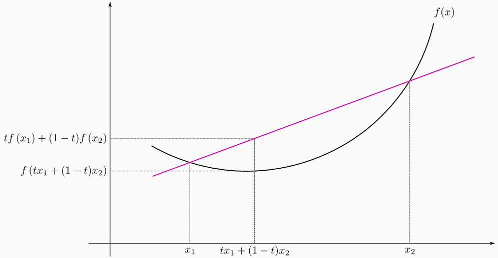

56

# Convexity and Jensen inequality (Optional)

We now introduce the notion of convex functions, one of most useful concept in applied mathematics.

**Definition**
Let $a < b$. A function $f\colon [a,b] \mapsto \mathbb{R}$ is convex if and only if, for all $x, y \in [a,b]$ and all $t \in [0,1]$,

$$f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$$



We first characterize differentiable convex functions

**Proposition**
*Let* $f : [a, b] \mapsto \mathbb{R}$ *be differentiable. Then* $f$ *is convex if and only if for all* $x, y \in [a, b]$

$$f(x) \geq f(y) + f'(y)(x - y)$$


58

We first characterize differentiable convex functions

**Proposition**
*Let $f : [a, b] \mapsto \mathbb{R}$ be differentiable. Then $f$ is convex if and only if for all $x, y \in [a, b]$*

$$f(x) \geq f(y) + f'(y)(x - y)$$

Proof First suppose that $f$ is convex. Then

$$f(x) \geq f(y) + \frac{f(tx + (1 - t)y) - f(y)}{t}$$

and hence, by taking the limit as $t \to 0$,

$$f(x) \geq f(y) + \lim_{t \to 0} \frac{f(y + t(x - y)) - f(y)}{t} = f(y) + f'(y)(x - y)$$



58

For the converse, suppose $f$ satisfy $f(x) \geq f(z) + f'(z)(x - z)$ for all $x, z$.

Let $x, y$ be given and $t \in [0, 1]$ be arbitrary. Then, for $z = tx + (1 - t)y$, we have

$$f(x) \geq f(z) + f'(z)(x - z), \quad (*)$$
$$f(y) \geq f(z) + f'(z)(y - z), \quad (**)$$

Multiplying $(*)$ by $t$ and $(**)$ by $(1 - t)$ and adding them together yield

$$tf(x) + (1 - t)f(y) \geq f(z)$$

as desired.


59

The following result follows from the previous proposition.

**Proposition**
*Let $f : [a, b]$ be twice differentiable. Then $f$ is convex on $[a, b]$ if and only if $f''(x) > 0$ for all $x \in [a, b]$.*


60

The following result follows from the previous proposition.

**Proposition**
*Let $f : [a, b]$ be twice differentiable. Then $f$ is convex on $[a, b]$ if and only if $f''(x) > 0$ for all $x \in [a, b]$.*

Proof We only show that $f''(x) > 0$ for all $x$ implies $f$ is convex.

This is a direct consequence of Taylor's series with remainder

$$f(x) = f(x_0)+f'(x_0)(x-x_0)+f''(x_*)(x-x_0)^2 \geq f(x_0)+f'(x_0)(x-x_0)$$

where $x_* \in [x_0, x]$. Thus $f$ is convex.

Example The following functions are convex

$$f(x) = x^2; \quad f(x) = e^x; \quad f(x) = 1/x; \quad f(x) = -\ln x$$


60

# Jensen's inequality

**Theorem**
*Let $f$ be a convex function. Then for all random variables $X$*

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$$

*provided that both expectations exist.*

Example Jensen's inequality implies

$$\mathbb{E}[X^2] \geq (\mathbb{E}[X])^2; \quad \mathbb{E}[e^{tX}] \geq \exp(t\mathbb{E}[X]); \quad \mathbb{E}[1/X] \geq 1/\mathbb{E}[X].$$


61

# Proof of Jensen's inequality

We only prove the result when $f$ is differentiable.

Let $\mu = \mathbb{E}[X]$. Then for all $x$

$$f(x) \geq f(\mu) + f'(\mu)(x - \mu) \Longrightarrow f(X) \geq f(\mu) + f'(\mu)(X - \mu)$$

We therefore have

$$\mathbb{E}[f(X)] \geq \mathbb{E}[f(\mu) + f'(\mu)(X - \mu)] = f(\mu) + f'(\mu)\mathbb{E}[X - \mu] = f(\mu)$$

as desired.


62