Refresher Course in Calculus, Probability, and Statistics

Day 3: Probability

Introduction

“Life’s most important questions are, for the most part, nothing but probability problems.“
(Pierre-Simon Laplace)

“Doubt is not a pleasant condition, but certainty is absurd.“
(Voltaire)

- Most aspects of the world around us have an element of randomness or uncertainty:
  - Will it rain tomorrow?
  - Will I win the lottery next week?
  - Will I earn more than 6,000 CHF/month some day?
- Theory of probability provides mathematical tools for quantifying and describing this randomness and dealing with uncertainty.
- Probability is needed to understand regression analysis and econometrics
- References:
  - [BWA] chap. 2; [SWA] chap. 2; [HGL] chap. 1 + appendix B

Probability
Basic definintions

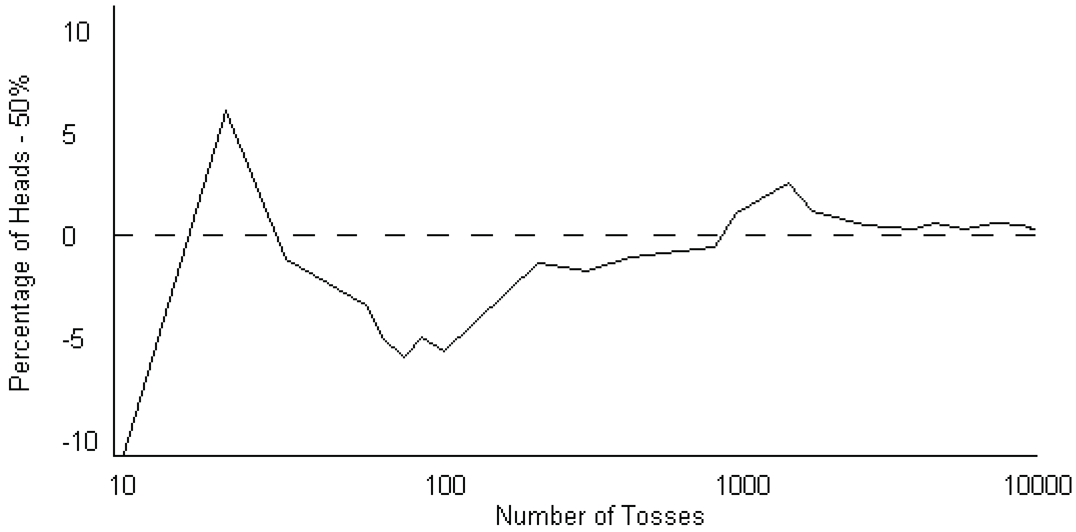
- **Trial**: event whose outcome is unknown (also called: experiment or observation)
  - Flipping a coin
  - Rolling a dice
- **Sample Space**: specification of all possible outcomes of a trial, denoted S
  - For flipping a coin the sample space is: $S = \{heads, tails\}$
  - For rolling a dice the sample space is: $S = \{1, 2, 3, 4, 5, 6\}$
- **Outcome**: mutually exclusive potential results of a random process
- **Event**: specification of the outcome of one trial (single outcome or a set)
  - The event "heads in flipping a coin": $A = \{heads\}$
  - The event "odd number in rolling a dice": $B = \{1, 3, 5\}$
- **Mutual exclusivity**: events that cannot occur together are mutual exclusive (e.g. passing a test and not passing a test)
- **Independence**: the outcome of one trial has no relationship to the outcome of another trial

Probability
Basic definintions (continued)

➲ **Relative frequency definition of probability**:

If an experiment is repeated n times under essentially identical conditions and the event A occurs m times, then as n gets large the ratio $\frac{m}{n}$ approaches the probability of A.

$$P(A) = \lim_{n\to\infty} \frac{m}{n}$$



Kerrich's experiment: Tossing a coin 10,000 times

Probability

Properties

Event	Probability
A	$P(A) \in [0,1]$
not A	$P(A^c) = 1 - P(A)$
A or B	$P(A \cup B) = P(A) + P(B)$ if A and B are mutually exclusive $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
A and B	$P(A \cap B) = P(A)P(B)$ if A and B are independent $P(A \cap B) = P(A\|B)P(B) = P(B\|A)P(A)$
A given B	$P(A\|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B\|A)P(A)}{P(B)}$

Random Variables

- **Random variable** (or stochastic variable): variable whose values result from measurement of a random process
  - Numerical summary of a random process
  - Probabilities are associated with possible values of random variable

- Example of a random variable:
  - number of times word crashes while writing a term paper
  - number of times heads comes up when tossing a coin 10 times
  - length of time it takes a random student to undertstand Bayes's theorem

- Random variables can be:
  - **Discrete**: can take only a limited number of values (gender, toss of a coin, roll of a die, ...)
  - **Continuous**: can take any value in an interval (time, earnings, prices, ...)

Discrete Random Variables

Probability Density Function

- **Probability density function (*pdf*)**: summarizes probabilities of possible outcomes

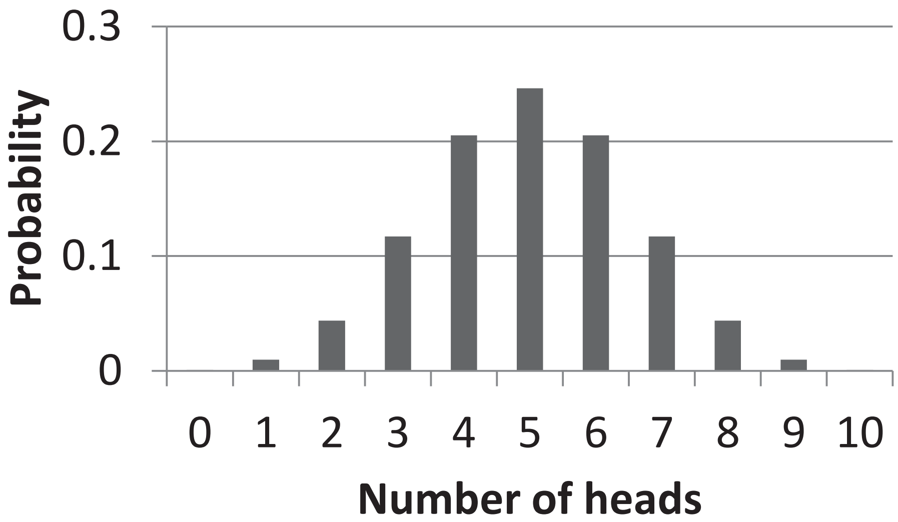
- For discrete random variables, it's like a table contrasting all possible values of the variable with the probability that each value will occur
  - indicates the probability of each possible value occuring
  - (the value of) *pdf* of discrete random variable X, denoted $f(x)$, is the probability that X takes value x:
$$f(x) = P(X = x)$$
  - $0 \leq f(x) \leq 1$
  - If X takes n possible values $x_1, \ldots, x_n$:
$$P(X = x_1) + \cdots + P(X = x_n) = f(x_1) + \cdots + f(x_n) = 1$$

Discrete Random Variables

Probability Density Function: Example

- *pdf* for number of heads out of 10 coin tosses:

	Outcome (number of heads)										
	0	1	2	3	4	5	6	7	8	9	10
$f(x)$	0.001	0.0098	0.0439	0.1172	0.2051	0.2461	0.2051	0.1172	0.0439	0.0098	0.001



Discrete Random Variables

Probability Density Function: Example (continued)

➲ *pdf* for number of heads out of 10 coin tosses (continued):

	Outcome (number of heads)										
	0	1	2	3	4	5	6	7	8	9	10
$f(x)$	0.0010	0.0098	0.0439	0.1172	0.2051	0.2461	0.2051	0.1172	0.0439	0.0098	0.0010

➲ Probability of an event (or compound event) can be computed from *pdf*

- Probability of event "exactly 7 heads":

$$P(X = 7) = 0.1172 = 11.72\%$$

- Probability of compound event "even number of heads"

$$P(X = 0) + P(X = 2) + P(X = 4) + P(X = 6) + P(X = 8) + P(X = 10)$$
$$= 0.0010 + 0.0439 + 0.2051 + 0.2051 + 0.0439 + 0.0010 = 0.5 = 50\%$$

- Probability of compound event "no more than 3 heads":

$$P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3)$$
$$= 0.001 + 0.0098 + 0.0439 + 0.1172 = 0.1719 = 17.19\%$$

Discrete Random Variables

Cumulative Density Function

- **Cumulative distribution function (*cdf*)**: alternative way to represent probabilities
- *cdf* of random variable X, denoted $F(x)$, is the probability that X is less than or equal to x:

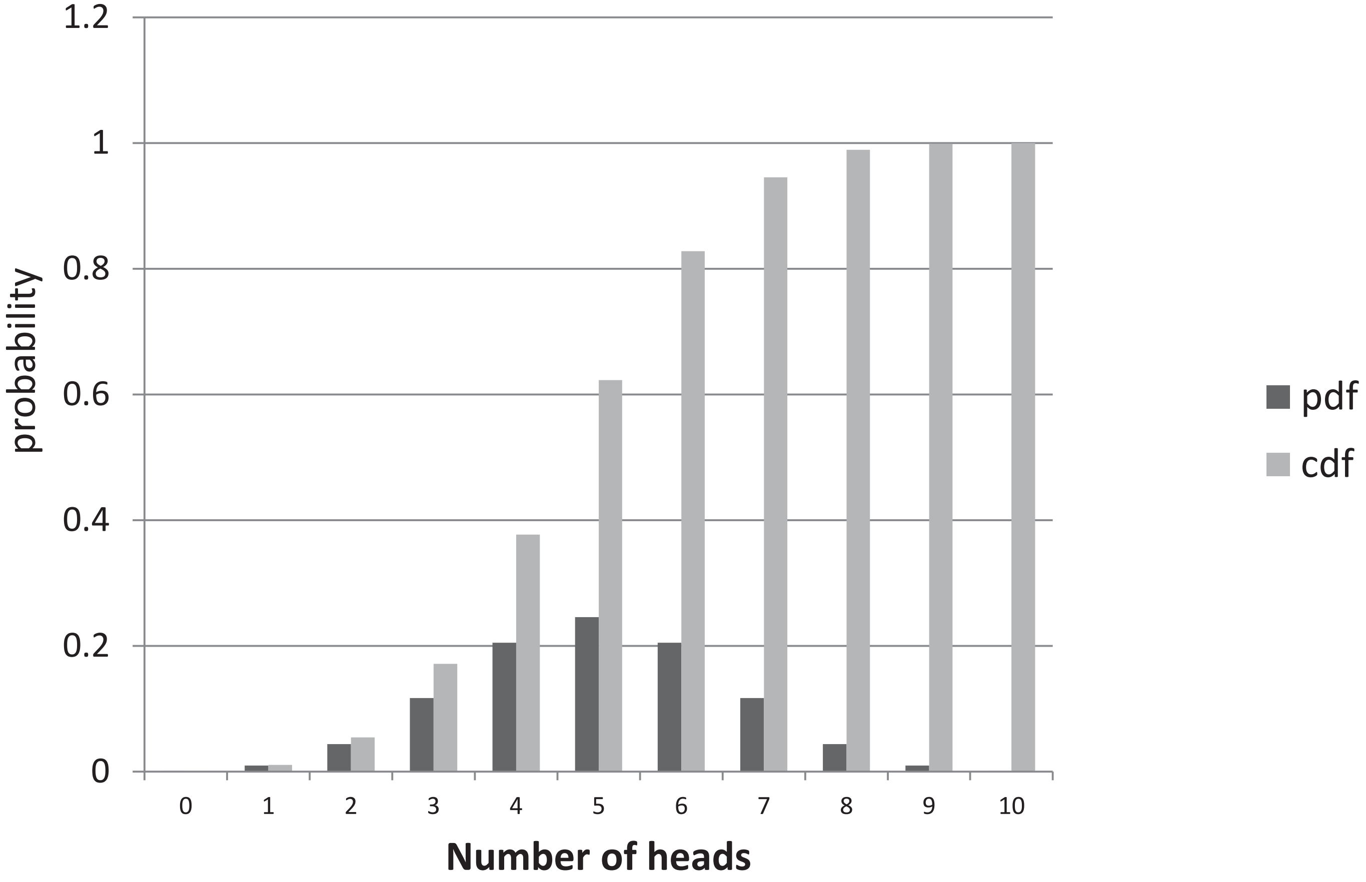
$$F(x) = P(X \leq x)$$

- *cdf* for number of heads out of 10 tosses:

	Outcome (number of heads)										
	0	1	2	3	4	5	6	7	8	9	10
$f(x)$	0.001	0.0098	0.0439	0.1172	0.2051	0.2461	0.2051	0.1172	0.0439	0.0098	0.001
$F(x)$	0.001	0.0108	0.0547	0.1719	0.3770	0.6231	0.8282	0.9454	0.9893	0.9990	1

Discrete Random Variables

Cumulative Density Function (continued):



Discrete Random Variables

Special Case: Binary Variables

- Important special case of discrete random variable: random variable that can take only two possible values (0 or 1)
  - Such variables are called binary variables, indicator variables, or dummy variables, dichotomous variables
- Probability distribution of binary random variable: **Binomial** or **Bernoulli distribution**
  - Sequence of independent Bernoulli trials n with constant probability of success at each trial p. We are interested in the total number of successes x.
- Example: In the 4^{th} quarter of 1988 in Massachusetts, USA, 60 new borns tested positive for HIV antibodies.
  - What are the chances that 30 children will be infected?
  - What are the chances that 30+ are infected?
  - Possible model: binomial with $p = 0.25$ and $n = 60$.

Jakob Bernoulli, 1655-1705

Discrete Random Variables
Special Case: Binary Variables (continued)

Binomial coefficient

- The Binomial distribution:

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

where x is the number of successes, p probability of success, and n number of trials (p and n are parameters of the model).

- Probability that 30 newborns will be HIV-positive:
  - o $P(30) = \binom{60}{30} 0.25^{30} (1-0.25)^{60-30} = \frac{60!}{30!(30)!} \cdot 0.25^{30} \cdot 0.75^{30} \approx 0.00018 = 0.018\%$

- So the probability that 30 or more of the newborns will be HIV-positive:

$$P(x \geq 30) = 1 - P(x < 30) = 1 - \sum_{i=0}^{29} P(x = i) \approx 0.00027 = 0.027\%$$

- A good visualization of the logic of the binomial distribution is the quincux: http://www.mathsisfun.com/data/quincunx.html

Discrete Random Variables
Expected Value

- **Mathematical expectation**: mean of random variable (long run average value of random variable over many repeated trials or occurrences)
- Expected value of discrete random variable X, taking values $x_1, \ldots, x_n$:

$$\begin{aligned} E(X) = \mu_X &= x_1 P(X = x_1) + \cdots + x_n P(X = x_n) \\ &= x_1 f(x_1) + \cdots + x_n f(x_n) \\ &= \sum_{i=1}^{n} x_i f(x_i) = \sum_{x} x f(x) \end{aligned}$$

  - weighted average of the possible outcomes (weights = probabilities)

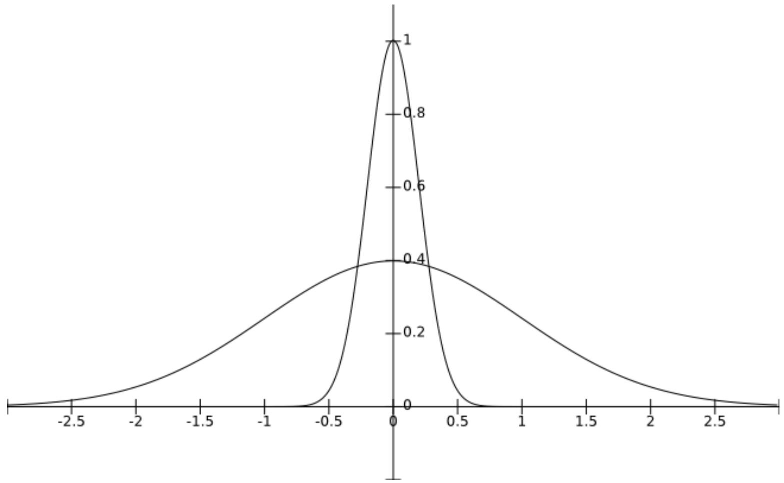
- For binary variables: $E(B) = 1 \times p + 0 \times (1 - p) = p$
  - For the binomial distribution with n trials: $E(B) = \mu_B = np$

- For linear combinations: $E(aX + bY + c) = aE(X) + bE(Y) + c$

Discrete Random Variables

Variance and Standard Deviation

- Variance and standard deviation measure dispersion or "spread" of probability distribution
  - Larger variance and standard deviation $\Longrightarrow$ more "spread out" values



- **Variance** of discrete random variable X:

$$var(X) = \sigma_X^2 = E[(X - \mu_X)^2] = E(X^2) - \mu_X^2$$

- For a binary variable: $var(B) = p(1 - p)$
- For the binomial distribution with n trials: $var(B) = np(1 - p)$
- For linear combinations: $var(aX + bY + \mathrm{c}) = a^2 var(X) + b^2 var(Y) + 2ab \cdot cov(X, Y) = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$

Discrete Random Variables
Variance and Standard Deviation (continued)

- **Standard deviation**:

$$\sigma_X = \sqrt{var(X)} = \sqrt{\sigma_X^2}$$

  - Same units of measure as the random variable

- For our example with the new borns:
  - How many should we expect to be HIV-positive?

    $$E(B) = \mu_B = np = 60 \times 0.25 = 15$$

  - Apply an empirical rule that $2/3$ of a (symmetric) distribution is covered in a range of $\mu_B \pm \sigma_X$ and 95% is covered by $\mu_B \pm 2\sigma_X$

    $$\sigma_B = \sqrt{var(B)} = \sqrt{np(1-p)} = \sqrt{60 \times 0.25(1-0.25)} = \sqrt{11.25} \approx 3.3541$$

  - $15 \pm 3.3541 = (11.65, 18.35) \approx [11, 19]$
  - $15 \pm 6.7082 = (8.29, 21.71) \approx [8, 22]$
  - Note: in our case the distribution is not perfectly symmetric
    - More exact intervals are $[11.56, 18.99]$ and $[8.83,\ 22.72]$

Discrete Random Variables
Other Measures of the Shape of a Distribution

- **Skewness** measures lack of symmetry:

$$Skewness = \frac{E[(X - \mu_X)^3]}{\sigma_X^3}$$

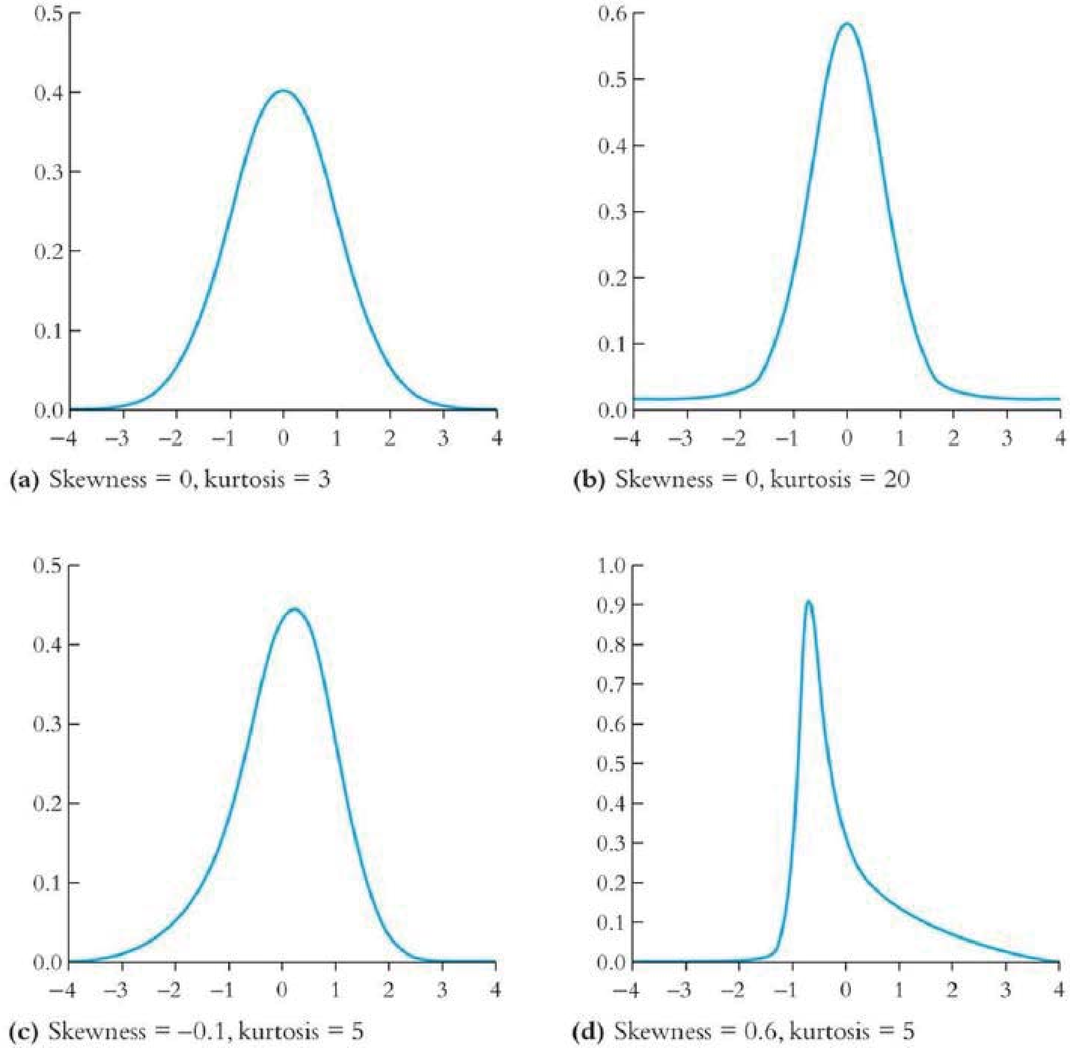
  - $S = 0$: symmetric, $S < 0$: left-skewed (long left tail), $S > 0$: right-skewed (long right tail)
  - Skewness is unitless

- **Kurtosis** measures how much mass in the tails of distribution:

$$Kurtosis = \frac{E[(X - \mu_X)^4]}{\sigma_X^4}$$

  - How much of the variance of X comes from extreme values (called: outliers)
  - Benchmark value is 3 (normal distribution); $K - 3$ called "*excess kurtosis*"
    - $K = 3$: mesokurtic, $K > 3$: leptokurtic, $K < 3$: platykurtic
  - Kurtosis is unitless and cannot be negative

Four Distributions with Different Skewness and Kurtosis



Stock & Watson (2007): 68

Several Discrete Random Variables

Joint and Marginal Distributions

- **Joint probability** that $X = x$ and $Y = y$ (joint *pdf* of X and Y)

$$f(x, y) = P(X = x, Y = y)$$

  - Sum of all joint probabilities: $\sum_x \sum_y f(x, y) = 1$

- **Marginal probability distribution** of random variable X:

$$f_X(x) = \sum_y f(x, y) \qquad for\ each\ value\ of\ X$$

  - Just another name for its probability distribution
  - Computed from joint distribution of X and Y by adding up probabilities of all possible outcomes for which X takes on a specific value

Several Discrete Random Variables
Conditional Distribution

- **Conditional *pdf*** : probability that random variable Y takes value y *given that* $X = x$:

$$P(Y = y|X = x) = \frac{P(X = x, Y = y)}{P(X = x)}$$

$$f(y|x) = \frac{f(x, y)}{f_X(x)}$$

- Statistical independence:

$$P(Y = y|X = x) = P(Y = y)$$

$$f(y|x) = f(y) = f_Y(y)$$

Several Discrete Random Variables
Example

- Joint distribution of computer crashes and operating system

Operating system	$M=0$	$M=1$	$M=2$	$M=3$	$M=4$	**Total**
Windows ($OS=0$)	0.531	0.08	0.05	0.029	0.01	0.7
Linux ($OS=1$)	0.269	0.02	0.01	0.001	0.00	0.3
Total	0.80	0.10	0.06	0.03	0.01	1.00

- Probability that computer will not crash at all and be running on Linux:
  - Joint probability: $P(M=0, OS=1) = 0.269 = 26.9\%$
- Probability any random computer is running on Windows:
  - Marginal probability: $P(OS=0) = \sum_{i=0}^{4} P(M=x_i, OS=0) = 0.7 = 70\%$
- Probability that a computer that doesn't crash when running on Linux:
  - Conditional probability: $P(M=0|OS=1) = \frac{P(OS=1, M=0)}{P(OS=1)} = \frac{0.269}{0.3} \approx 0.8967 = 89.67\%$
- Expectation of probability of no crash with Linux given statistical independence:
  - $P(M=0|OS=1) = P(M=0) = 0.80$

Several Discrete Random Variables

Conditional Expectation

➲ Conditional expectation (also called: conditional mean) of Y given $X = x$, if Y can take on k values $y_1, \dots, y_k$:

$$E(Y|X = x) = \sum_{i=1}^{k} y_i P(Y = y_i|X = x) = \sum_{y} y f(y|x)$$

- Example: How many crashes do I expect when working with
  - Linux computer: $E(M|OS = 1) = \sum_{i=0}^{4} m_i P(M = m_i|OS = 1) = 0 \cdot \frac{0.269}{0.3} + 1 \cdot \frac{0.02}{0.3} + 2 \cdot \frac{0.01}{0.3} + 3 \cdot \frac{0.001}{0.3} + 4 \cdot \frac{0.00}{0.3} \approx 0.143$
  - Windows computer: $E(M|OS = 0) = \sum_{i=0}^{4} m_i P(M = m_i|OS = 0) = 0 \cdot \frac{0.531}{0.7} + 1 \cdot \frac{0.08}{0.7} + 2 \cdot \frac{0.05}{0.7} + 3 \cdot \frac{0.029}{0.7} + 4 \cdot \frac{0.01}{0.7} \approx 0.439$

Several Discrete Random Variables

Conditional Variance

➲ Conditional variance (variance of Y given a value for X):

$$var(Y|X=x) = \sum_{i=1}^{k} [y_i - E(Y|X=x)]^2 P(Y=y_i|X=x)$$

- Example: Conditional variance of crashes when working with:
  - Linux computer:

$$var(M|OS=1) = \sum_{i=1}^{4} \{[m_i - E(M|OS=1)]^2 P(M=m_i|OS=1)\}$$

$$= (0-0.143)^2 \cdot \frac{0.269}{0.3} + (1-0.143)^2 \cdot \frac{0.02}{0.3} + (2-0.143)^2 \cdot \frac{0.01}{0.3}$$

$$+ (3-0.143)^2 \cdot \frac{0.001}{0.3} + (4-0.143)^2 \cdot \frac{0.00}{0.3} \approx 0.209$$

  - Windows computer:

$$var(M|OS=1) = \sum_{i=0}^{4} \{[m_i - E(M|OS=1)]^2 P(M=m_i|OS=1)\}$$

$$\approx 0.809$$

Several Discrete Random Variables

Covariance and Correlation

➲ **Covariance** between X and Y:

$$cov(X, Y) = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - \mu_X\mu_Y$$

- $\sigma_{XY} > 0$: positive association (when $X > \mu_X$ then $Y > \mu_Y$ and when $X < \mu_X$ then $Y < \mu_Y$)
- $\sigma_{XY} < 0$: negative association
- $\sigma_{XY} = 0$: neither negative nor positive association

➲ **Correlation** between X and Y:

$$corr(X, Y) = \rho = \frac{cov(X, Y)}{\sqrt{var(X)}\sqrt{var(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

- Pearson product-moment correlation coefficient
- unitless measure
- $-1 \leq \rho \leq 1$

Karl Pearson, 1857-1936

Continuous Random Variables

- Continuous random variables can take any value in an interval
  - GDP, interest rates, income, stock market indices, exchange rates, ... are (treated as) continuous variables

- Because they can take uncountable many values, probability that any single value occurs is zero
  - For example: if X is tomorrow's *EUR*|*CHF* exchange rate, $P(X = 1.240) = 0$

- For continuous variables, probability statements are meaningful when we consider outcomes within intervals
  - For example: if X is tomorrow's *EUR*|*CHF* exchange rate, $P(X = 1.240) \geq 0$ (because 1.2400 is usually considered an interval, e.g. $[1.2395, 1.2404[$)

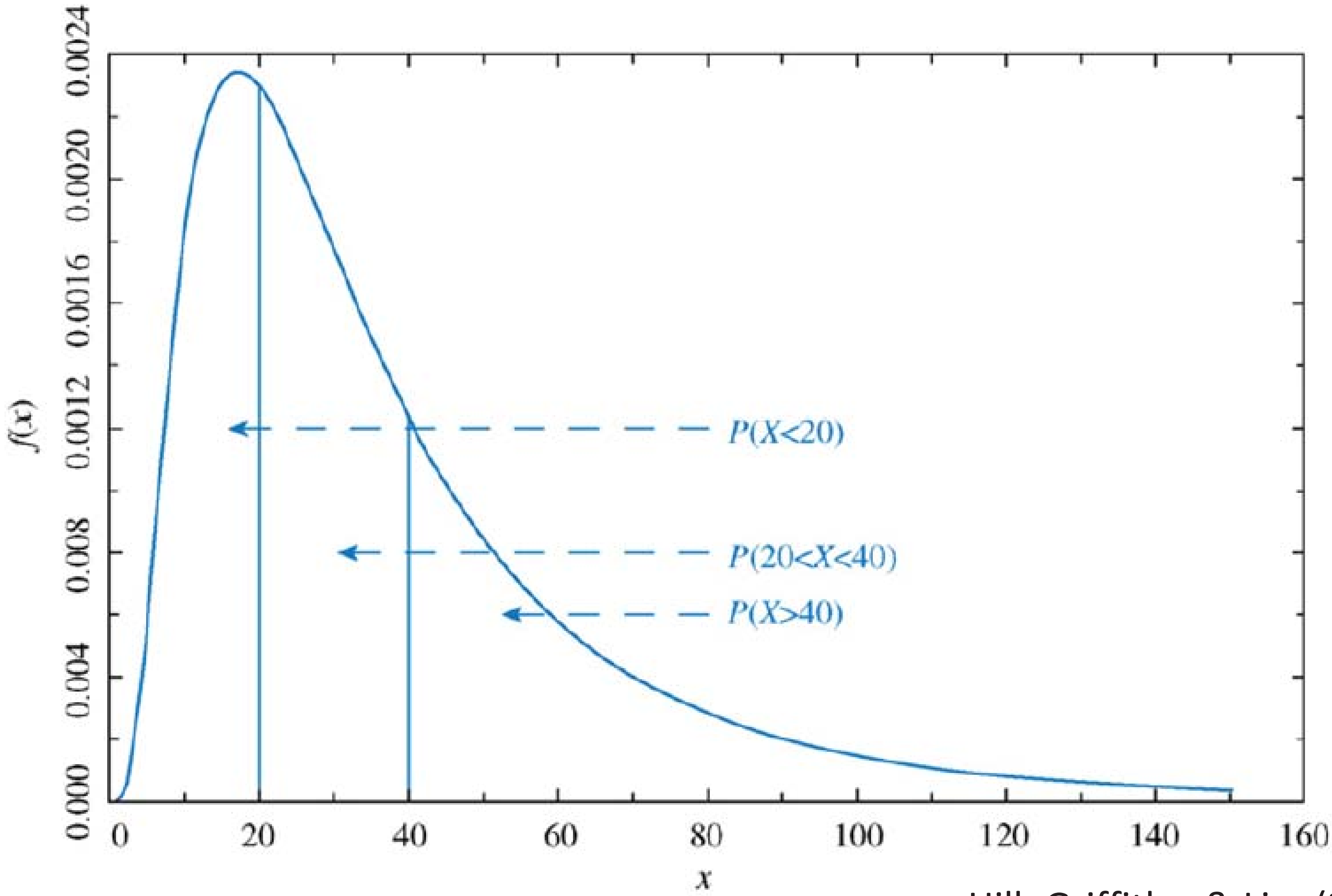
Continuous Random Variables

Probability Density Function

- **Probability density function (*pdf*)** summarizes the probability for a continuous variable
- Let X be a continuous random variable with *pdf* $f(x)$
  - $f(x) \geq 0$
  - $\int_{-\infty}^{\infty} f(x)dx = 1$
  - $P(a \leq X \leq b) = \int_{a}^{b} f(x)dx$

    Probability that X falls in the interval $[a, b]$ = the area under $f(x)$ between these points

pdf of a Continuous Random Variable



Hill, Griffiths, & Lim (2008)

Continuous Random Variables

Cumulative Density Function

- **Cumulative probability distribution (*cdf*)** defined as for discrete variable: probability that random variable is less than or equal to a specific value

$$P(X \leq a) = \int_{-\infty}^{a} f(x)dx = F(a)$$

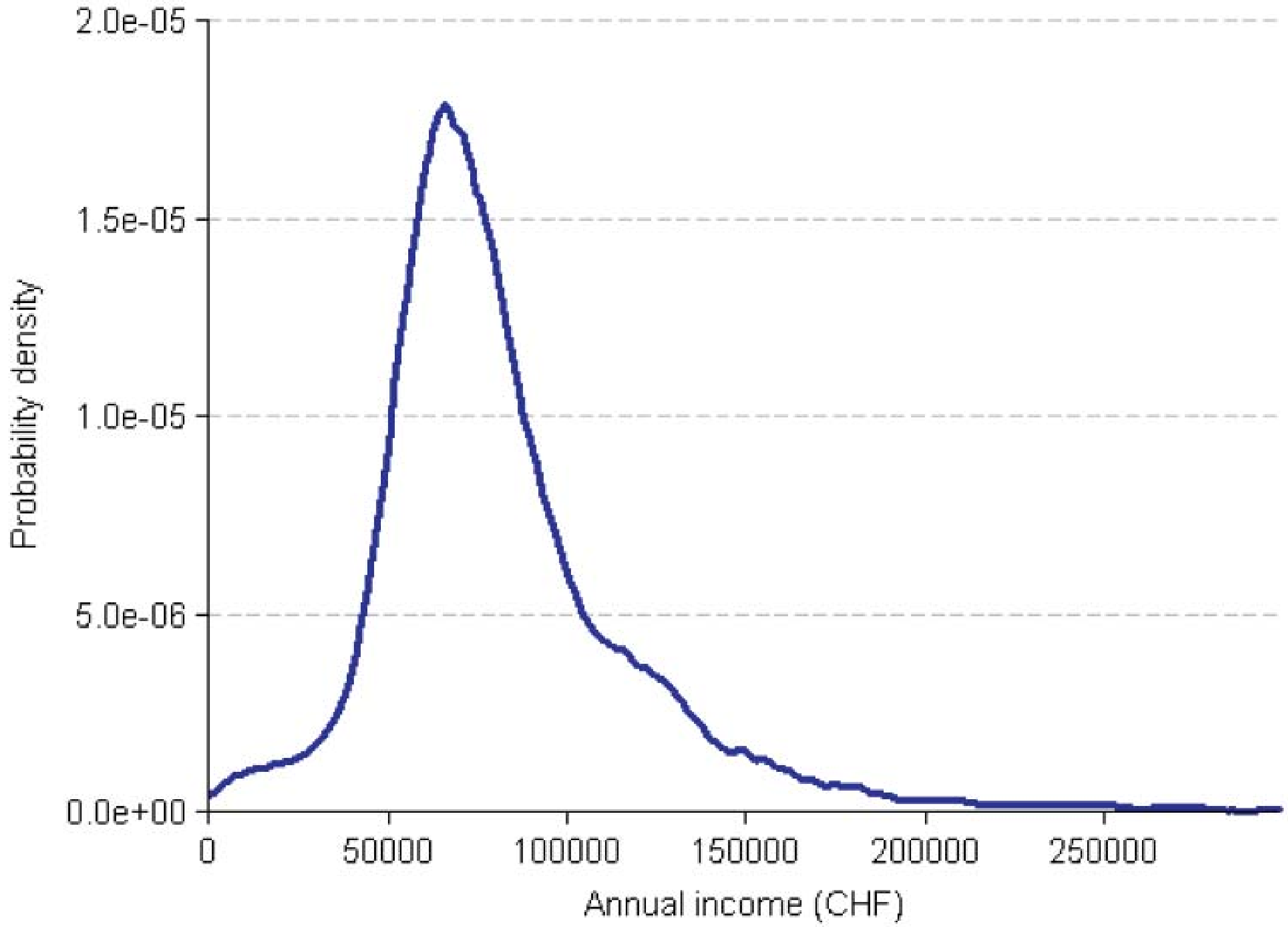
- *cdf* obtained by integrating *pdf*

- Hence, Possible to obtain *pdf* by differentiating *cdf*:

$$f(x) = \frac{dF(x)}{dx} = F'(x)$$

An Empirical *pdf*

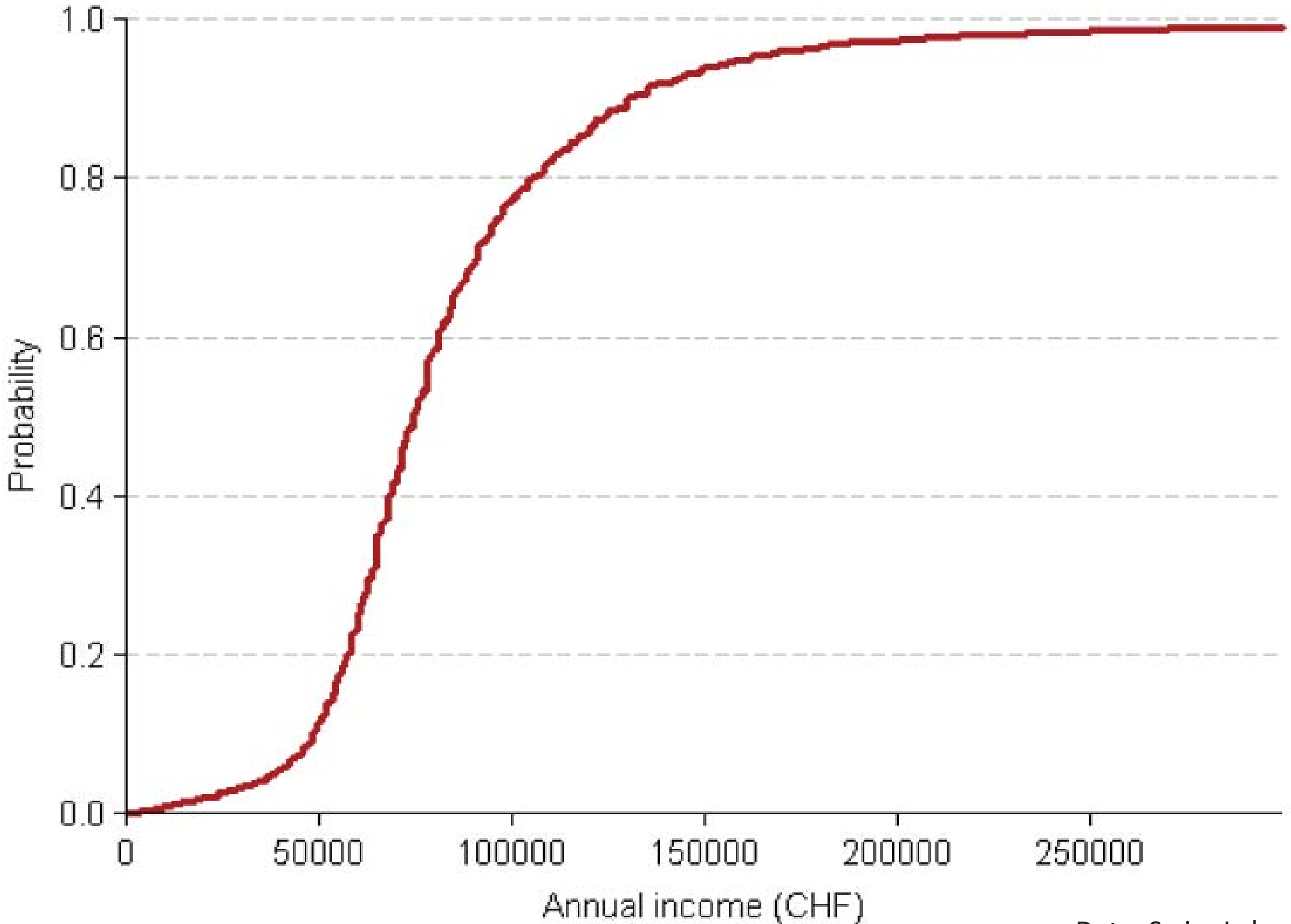
Male Workers' Income in Switzerland



Data: Swiss Labor Force Survey 2008

An Empirical *cdf*

Male Workers' Income in Switzerland



Data: Swiss Labor Force Survey 2008

Continuous Random Variables

Expectation and Variance

- Expectation:

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

  - Compared to discrete case, integral simply replaces summation
  - Interpretation: average value of X for an infinite number of repetitions

- Variance:

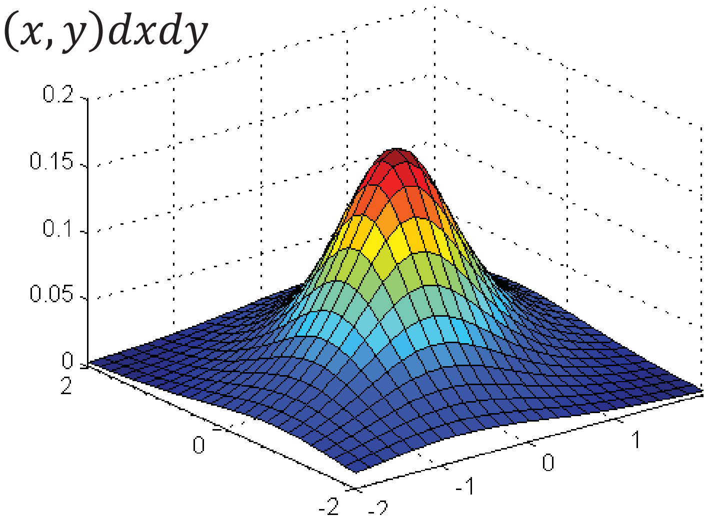
$$\sigma_X^2 = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x) dx = E(X^2) - \mu_X^2$$

Several Continuous Random Variables

Joint, Marginal, and Conditional Probability Distributions

- Joint probability density function $f(x, y)$ is a surface and probabilities are volumes under the surface
- Probability that X is between a and b and at the same time Y is between c and d:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_{x=a}^{b} \int_{y=c}^{d} f(x, y)dxdy$$



- Marginal probability density function for Y:

$$f(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

- Conditional probability density function:

$$f(y|x) = \frac{f(x, y)}{f(x)}$$

  - Unlike the discrete case $f(x|y)$ is not a probability but a density function that can be used to compute probabilities:

$$E(Y|X = x) = \int_{-\infty}^{\infty} yf(y|x)dy$$

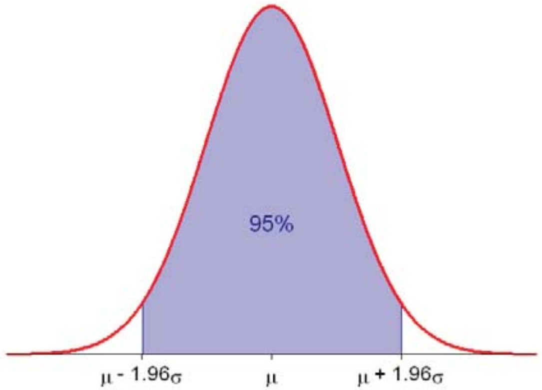
Normal Distribution

- **Normal distribution (or Gaussian distribution)**: continuous probability distribution with bell-shaped *pdf*
- Normally distributed random variable X with mean μ and variance σ^2 symbolized:

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

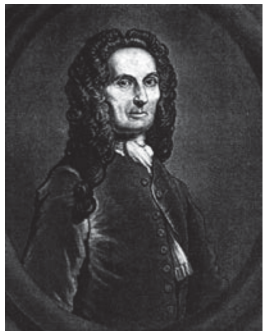
- *pdf* of X:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}$$



- Standard normal distribution: $Z \sim N(0,1\ )$
  - Computer softwares and tables provide values of standard normal *cdf*, usually denoted $\Phi(z) = P(Z \leq z)$
  - **Standardization** to compute probabilities for variable $X \sim N(\mu, \sigma^2)$:

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



Abraham de Moivre, 1667-1754

Normal Distribution (continued)

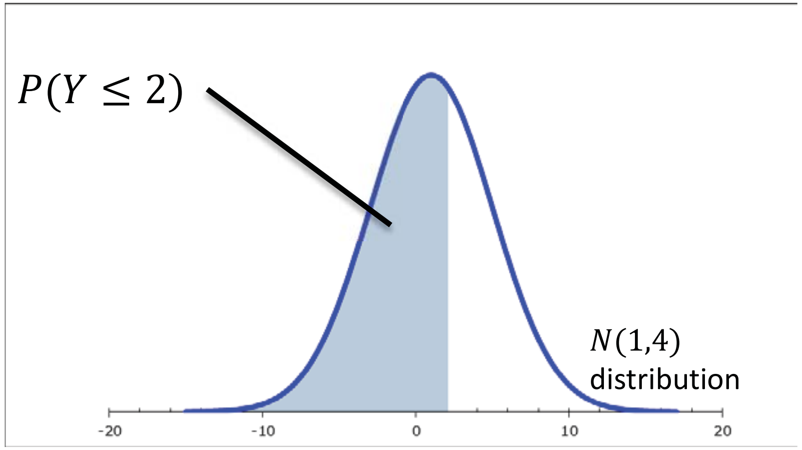
Calculating probabilities

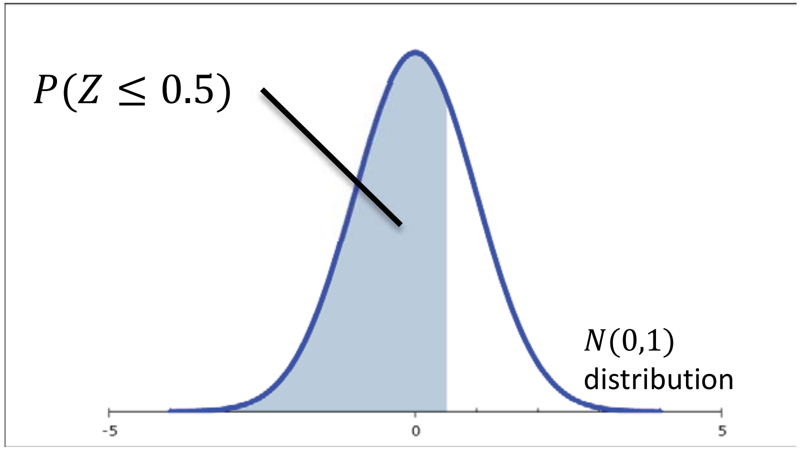
- What is the probability that $Y \leq 2$ if $Y \sim N(1,4)$
  - Step 1: z-standardize to get z-score

$$z = \frac{x - \mu}{\sigma} = \frac{2 - 1}{2} = 0.5$$

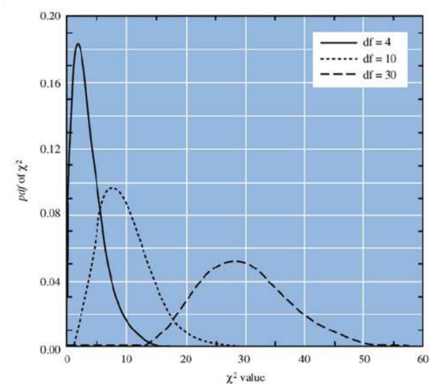
  - Step 2: compare z-score to distribution table

$$P\left(\frac{Y - 1}{2} \leq 0.5\right) = \phi(0.5) = 0.691$$

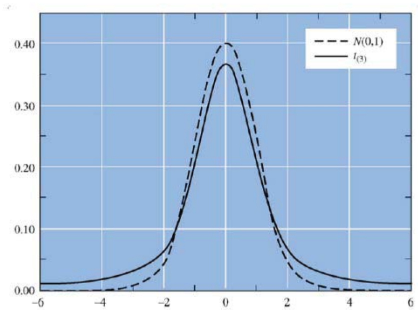




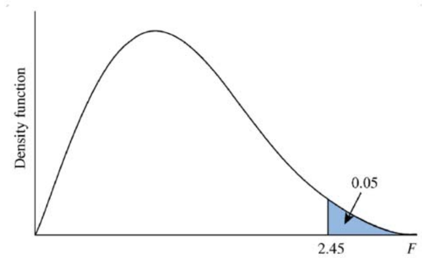
Other important distributions



- **Chi-square distribution**
  - o $V = Z_1^2 + Z_2^2 + \cdots + Z_m^2 \sim \chi^2_{(m)}$ with m degrees of freedom
  - o where $Z_i \sim N(0,1)$, $i = 1, 2, \ldots, m$ and Z_i are independent
    - As $m \to \infty$, $\chi^2_{(m)} \to Z \sim N(0,1)$ and $\chi^2_{(1)} = Z_1^2$



- **Student t-distribution**
  - o $t = \frac{Z}{\sqrt{\frac{V}{m}}} \sim t_{(m)}$ with m degrees of freedom
  - o where $Z \sim N(0,1)$ and $V \sim \chi^2_{(m)}$
    - As $m \to \infty, t_{(m)} \to Z \sim N(0,1)$



- **F-distribution**
  - o $F = \frac{\frac{V_m}{m}}{\frac{V_n}{n}} = \frac{\chi^2_m}{m} \cdot \frac{n}{\chi^2_n} \sim F_{(m,n)}$ with m, n degrees of freedom
  - o Relationship to Chi-square distribution: $\lim_{n\to\infty} F = \frac{\chi^2_m}{m}$