Probability and statistics for the FE Exam

General comments

1. FE Reference Manual pp. 33–49.
   — Don't be intimidated by the manual.
2. This is a big topic, but only a handful of concepts will get you a long way.
3. Work lots of sample problems.
4. Familiarize yourself with the contents of the ref. man.
5. There will be 4–6 prob. + stats questions on the exam. You will have an average of three minutes/problem.

A quick and dirty overview

Sets

A set is a collection of objects. Set theory gives us a way to describe these collections. Often, the objects in a set are numbers or sets of numbers. However, a set could represent collections of zebras and trees and hairballs.

Examples of sets: $\{1, 5, \pi\}$ | {zebra named "Calvin", a burnt cheeto} | $\{\{1,2\}, \{5, \text{hippo}, 7\}, 62\}$

The union of sets is a set containing all the elements in the original sets (sets don't have repeating elements, though). The union of sets A and B is denoted $A \cup B$.

Example of set union: Let $A = \{1, 2, 3\}$ and $B = \{-1, 6\}$. Then

$$A \cup B = \{-1, 1, 2, 6, 3\}.$$

The intersection of sets is a set containing the element common to all the original sets. The intersection of sets A and B is denoted $A \cap B$.

Example of set intersection: Let $A=\{1,2,3\}$ and $B=\{1,5,2\}$. Then

$$A \cap B = \{1,2\}$$

If two sets have no elements in common, the intersection is the empty set $\emptyset = \{\}$, the unique set with no elements.

The set difference of two sets A and B is the set of elements in A that aren't also in B. It's denoted $A \setminus B$.

Example of set difference: Let $A=\{1,2,3\}$ and $B=\{2,3,4\}$. Then

$$A \setminus B = \{1\} \quad \text{and} \quad B \setminus A = \{4\}.$$

A subset of a set is a set, the elements of which are all contained in the original set. If B is a subset of A, we can express this relation by $B \subseteq A$.

Example of a subset: Let $A=\{1,2,3\}$ and $B=\{1,2\}$. Then

$$B \subseteq A.$$

It is true that $A \subseteq A$: a set is a subset of itself.

The complement of a subset of a set is a set of the elements of the original set that aren't in the subset.

Example of a complement: Let A be a subset of B. Then the complement of A is $B \setminus A$

Combinations

The number of combinations of n set elements taken r at a time is denoted $C(n,r)$, C^n_r, ${}_nC_r$, or $\binom{n}{r}$.

We use the following equation to compute the number of possible combinations:

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} \qquad (r \leq n).$$

Permutations

The number of permutations of a set of n elements taken r at a time is denoted $P(n,r)$, P_r^n, or ${}_nP_r$ and is computed by the formula

$$P(n,r) = \frac{n!}{(n-r)!} \qquad (r \leq n).$$

Sample space and events

The sample space Ω of an experiment is the set representing all possible outcomes of the experiment.

If a coin is flipped, the sample space is $\Omega = \{\overset{\text{heads}}{H}, \overset{\text{tails}}{T}\}$.

If a coin is flipped twice, we can choose $\Omega = \{HH, HT, TH, TT\}$.

However, the same experiment can have different sample spaces. For instance for two coin flips, we could choose

$\Omega = \{$the flips are the same, the flips are different$\}$.

We base our choice of Ω on the problem at hand.

An event is a subset of the sample space. That is, an event corresponds to a yes-or-no question about the experiment. For instance, event A (remember: $A \subseteq \Omega$) in the coin flipping experiment (two flips) might be $A = \{HT, TH\}$. A is an event that corresponds to the question, "Is the second flip different than the first?" A is the event for which the answer is "yes."

Basic probability

The probability $P(A)$ of an event A is a real number in the interval $[0,1]$. The meaning of probability is not yet clearly understood, but we often use it to represent our confidence that an event will occur, and sometimes simply

$$P(A) = \frac{\#\text{ of outcomes in event } A}{\#\text{ of outcomes possible}}.$$

The probability of an event occuring plus the probability of it not occuring must equal 1:

$$P(A) + P(\text{not } A) = 1.$$

$\llcorner A', \neg A$

$P(A \cap B)$ // $P(A \cup$

Law of total probability:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Conditional probability

If events A and B are somehow dependent, we need a way to compute the probability of B occurring given that A has already occurred. This is called the conditional probability of B given A, and is denoted $P(B|A)$. It is defined as

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (\text{assuming } P(A) > 0).$$

Given two mutually exclusive and exhaustive events A and B, Bayes' Theorem states that

$$P(A|B) = P(B|A)\frac{P(A)}{P(B)}.$$

Sometimes this is written:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A)P(A) + P(B|A')P(A')}.$$

This is a useful theorem for determining a test's effectiveness.

If a test is performed to determine whether an event has occurred, we might ask questions like "If the test indicates that the event has occurred, what is the probability it actually has occurred? Bayes' rule can help compute an answer.

Four types of outcomes can occur:

1. True positive: test indicates occurrence and actual occurrence.
2. False positive: test indicates occurrence but no actual occurrence.
3. False negative: test indicates no occurrence but actual occurrence.
4. True negative: test indicates no occurrence and no actual occurrence.

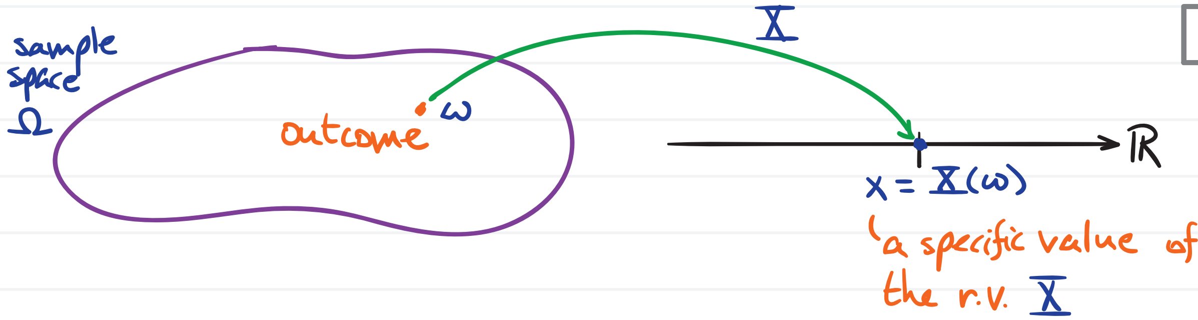
	True ☺	False ☹
Positive	1	2
Negative	4	3

As measurement engineers, we want to live in the first column as much as possible. Good tests yield 1 and 4 most of the time. Poor tests yield 2 and 3 too often.

Random variables

Probilities are useful even when they do not deal strictly with events. It often occurs that we measure something that has randomness associated with it. We use random variables to represent these measurements.

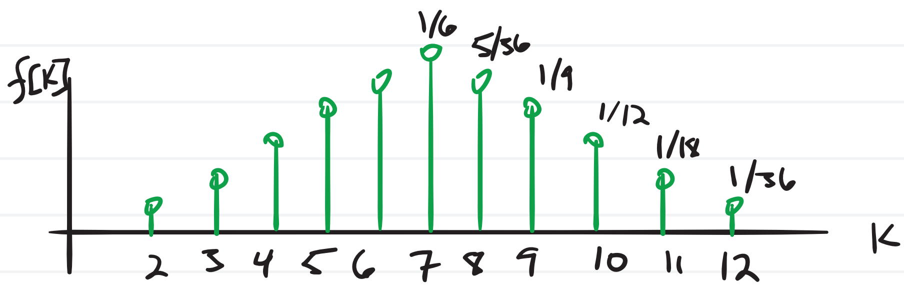
A random variable is a function that maps from the sample space to the real numbers, $X: \Omega \to \mathbb{R}$.



Discrete random variables

Discrete random variables take on discrete values.

Example Roll two unbiased dice. Let K be a r.v. representing the sum of the two. Draw the results.



Continuous random variables

Just like discrete, but take on values in a continuum.

Probability density functions

We typically think of a probability density function as a function that can be integrated over to find the probability of the random variable being in an interval:

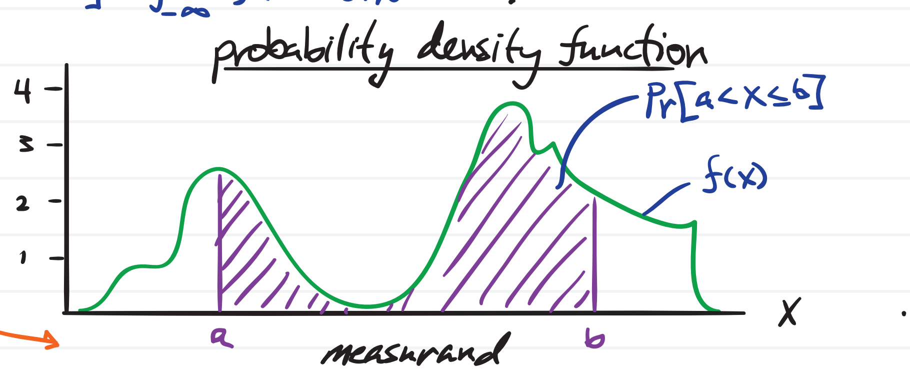
$$\Pr[a < x \le b] = \int_a^b f(x)\,dx .$$

Of course, $\Pr[-\infty < x < \infty] = \int_{-\infty}^{\infty} f(x)\,dx = 1.$

For instance,

can be >1

can never be <0



Binomial PMF

Consider a binary sequence of length n with each element a random 0 or 1 generated independently, like

$$1 \quad 0 \quad 1 \quad 1 \quad 0 \quad \cdots \quad 0 \quad 1 .$$

This sequence has the probability of occurring:

$$\overbrace{p}^{\text{prob. of a 1}} \cdot (1-p) \cdot p \cdot p \cdot (1-p) \cdots (1-p) \cdot p$$

$$= p^{\overbrace{k}^{\text{number of 1s.}}} (1-p)^{(n-k)}$$

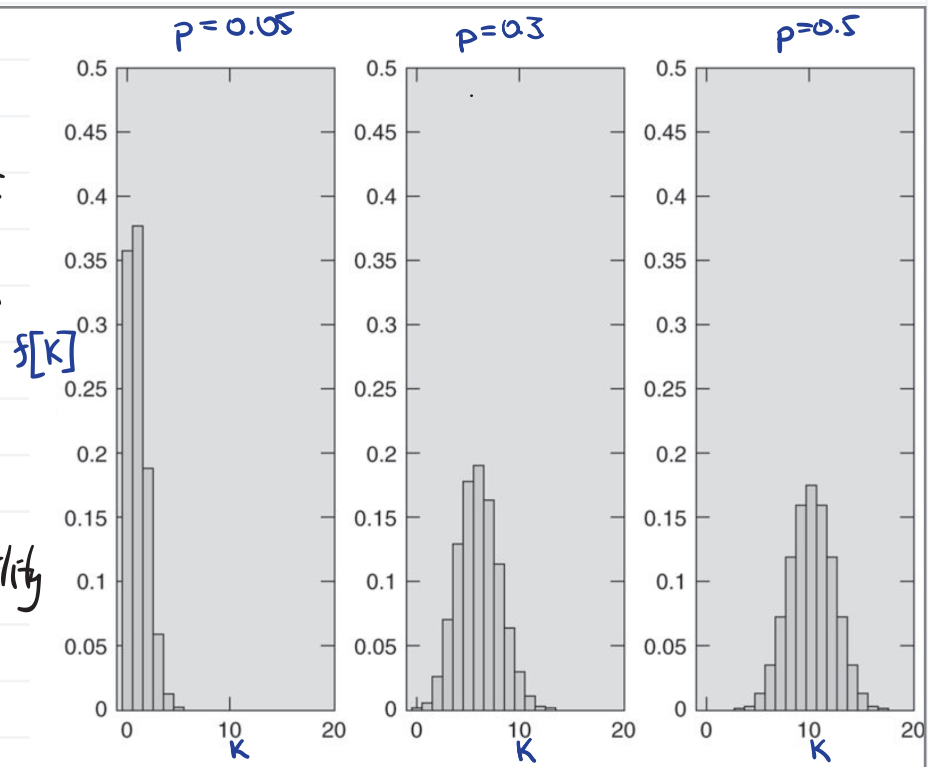
There are $\binom{n}{k}$ $\left(n \text{ choose } k = \frac{n!}{k!(n-k)!}\right)$ possible combinations of k 1s in n bits. Therefore, the probability of any combination of k 1s in a series is:

$$f[k] = \binom{n}{k} p^k (1-p)^{n-k} ,$$

which is the binomial distribution PMF.

Example Consider a field sensor that fails for a given measurement with probability p. Given n measurements, we can plot the Binomial PMF for different probabilities of failure (right).

These tell us the probability of getting k failures of the sensor in $n = 20$ measurements.



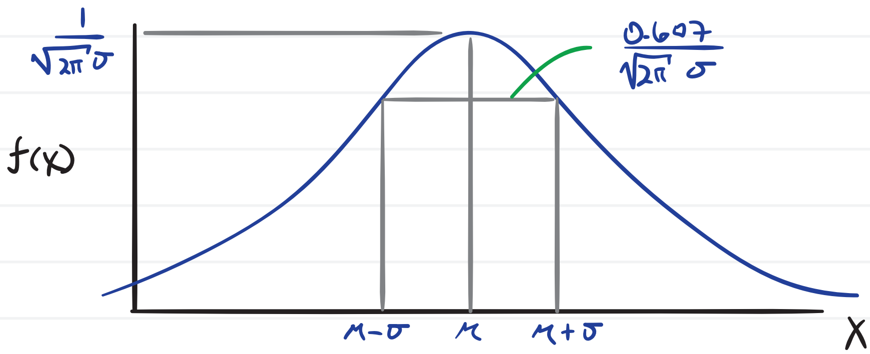
Gaussian PDF ("Normal" distribution)

The Gaussian or normal random variable has PDF:

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

where:
μ is the mean
σ is the standard deviation
σ^2 is the variance

This has the familiar "bell" shape:



It can be shown that $\Pr[\mu-\sigma < X \leq \mu+\sigma] = 68.2\%$.

Noise is often modelled as Gaussian. Many random processes are nearly Gaussian, so it is important to understand.

Student's t-distribution

Student's t-distribution is a PDF that is similar to the Gaussian PDF, but is often a better model for small sample sizes.

The r.v. $t_{\alpha,\nu}$ gives the number of $\pm$ standard deviations from the mean a precision interval predicts the next value of a r.v. with confidence $(1-2\alpha)$ (e.g. $\alpha = 0.05 \rightarrow$ confidence $1-2\cdot 0.05 = 90\%$).

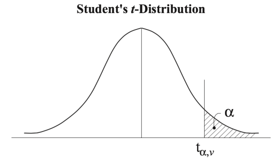
The FE Reference table (p. 43) is given in α and ν.

Example: Sample size $N=4$, sample mean $\bar{x}$. What is the 50% confidence precision interval for the next measurement? For μ? The sample std. dev. is S_x.

Ans: $\nu = N-1 = 3$ | $\alpha = \frac{1-0.50}{2} = 0.25$ | $x_i = \bar{x} \pm 0.765 \cdot S_x$.

Ans: $SDOM = S_x/\sqrt{N} = S_x/2$ | $\mu = \bar{x} \pm 0.765 \cdot S_x/2$.

Student's t-Distribution



VALUES OF $t_{\alpha,\nu}$

ν	0.25	0.20	0.15	0.10	0.05	0.025	0.01	0.005	ν
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	1
2	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925	2
3	0.765	0.978	1.350	1.638	2.353	3.182	4.541	5.841	3
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	4
5	**0.727**	**0.920**	**1.156**	**1.476**	**2.015**	**2.571**	**3.365**	**4.032**	**5**
6	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707	6

Let ν be the number of statistical degrees of freedom. This is defined as the number of samples N minus the number of constraints c ($\nu \equiv N - c$). When computing means, $c=0$. When computing standard deviations and variations we require the mean, so $c=1$.

Let Γ be the gamma function. Γ has the following properties:

$$\Gamma(n) = (n-1)! \quad \text{for } n \in \{\text{whole integer}\}$$
$$\Gamma(n) = (n-1)(n-2)\cdots(3/2)(1/2)\sqrt{\pi} \quad \text{for } m \in \{\text{half integers}\}$$
$$\Gamma(1/2) = \sqrt{\pi}\,.$$

The Student's t-distribution (PDF) is defined as

$$f(t,\nu) = \frac{\Gamma((\nu+1)/2)}{\sqrt{\pi\nu}\,\Gamma(\nu/2)}\left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}.$$

Note: actually a family of PDFs.

We usually look up t-values in a table to find precision/confidence intervals, as shown later.

Basic statistics

The arithmetic mean of a sample (sample mean) of a measurand represented by random variable X is defined as

$$\bar{x} \equiv \frac{1}{N}\sum_{i=1}^{N} x_i \,.$$

sample size

If the sample size is large, $\bar{x} \to \mu$ (the sample mean approaches the "true mean" or population mean). The population mean (or mean) can be expressed as

$$\mu = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} x_i \,.$$

Recall that we defined the mean as $\mu \equiv \langle X \rangle$.

The sample variance of a measurand represented by r.v. X is defined as

$$S_X^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \,.$$

If the sample size is large, $S_{\bar{X}}^2 \longrightarrow \sigma_{\bar{X}}^2$ (the sample variance approaches the "true variance" or population variance). The population variance can be expressed as

$$\sigma_{\bar{X}}^2 = \lim_{N\to\infty} \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2 .$$

Recall that we defined the variance as $\sigma_{\bar{X}}^2 = \langle (x-\mu)^2 \rangle$.

Confidence/precision intervals

It can be said that a normally distributed r.v. value x_i in a small sample will be within $\pm t_{\nu,P}$ sample standard deviations of the sample mean with $P\%$ confidence. That is,

$$x_i = \bar{x} \pm \underbrace{t_{\nu,P} S_x}_{\text{called the precision interval}} \quad (P\% \text{ confidence}) .$$

Example

2. Law of total probability: $A = \frac{1}{3}$ guy gets it right
$B = \frac{3}{4}$ guy gets it right

$$P(A) = \tfrac{1}{3} \qquad P(B) = \tfrac{3}{4}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$= \tfrac{1}{3} + \tfrac{3}{4} - \tfrac{1}{3} \cdot \tfrac{3}{4}$$

$$= \tfrac{13}{12} - \tfrac{3}{12} = \tfrac{10}{12} = \tfrac{5}{6}$$

1. A = three heads $\quad n = 4, \; k = 3$

$$\left.\begin{aligned} \#\text{ of 3-heads outcomes} &= \binom{4}{3} = \frac{4!}{3!(4-3)!} = 4 \\ \#\text{ of possible outcomes} &= 2^4 = 16 \end{aligned}\right\} \frac{4}{16} = \frac{1}{4}$$