Simulation

Dr. Xueping Li
University of Tennessee

A Refresher on Probability and Statistics

Appendix C

What We'll Do ...

- **Ground-up review of probability and statistics necessary to do and understand simulation**
- **Assume familiarity with**
  - Algebraic manipulations
  - Summation notation
  - Some calculus ideas (especially integrals)
- **Outline**
  - Probability – basic ideas, terminology
  - Random variables, joint distributions
  - Sampling
  - Statistical inference – point estimation, confidence intervals, hypothesis testing

Probability Basics

- ***Experiment*** **– activity with uncertain outcome**
  - Flip coins, throw dice, pick cards, draw balls from urn, …
  - Drive to work tomorrow – Time? Accident?
  - Operate a (*real*) call center – Number of calls? Average customer hold time? Number of customers getting busy signal?
  - *Simulate* a call center – same questions as above
- ***Sample space*** **– complete list of all possible individual outcomes of an experiment**
  - Could be easy or hard to characterize
  - May not be necessary to characterize

Probability Basics (cont'd.)

- ***Event*** **– a subset of the sample space**
  - Describe by either listing outcomes, "physical" description, or mathematical description
  - Usually denote by E, F, E_1, E_2, etc.
  - Union, intersection, complementation operations
- ***Probability*** **of an event is the relative likelihood that it will occur when you do the experiment**
  - A real number between 0 and 1 (inclusively)
  - Denote by $P(E)$, $P(E \cap F)$, etc.
  - Interpretation – proportion of time the event occurs in many independent repetitions (replications) of the experiment
  - May or may not be able to derive a probability

Probability Basics (cont'd.)

- **Some properties of probabilities**

  If S is the sample space, then $P(S) = 1$

    Can have event $E \neq S$ with $P(E) = 1$

  If Ø is the empty event (empty set), then $P(\varnothing) = 0$

    Can have event $E \neq \varnothing$ with $P(E) = 0$

  If E^C is the complement of E, then $P(E^C) = 1 - P(E)$

  $P(E \cup F) = P(E) + P(F) - P(E \cap F)$

  If E and F are mutually exclusive (i.e., $E \cap F = \varnothing$), then $P(E \cup F) = P(E) + P(F)$

  If E is a subset of F (i.e., the occurrence of E implies the occurrence of F), then $P(E) \leq P(F)$

  If o_1, o_2, ... are the individual outcomes in the sample space, then $\sum_{\text{all } i} P(o_i) = 1$

Probability Basics (cont' d.)

- ***Conditional*** **probability**
  - Knowing that an event F occurred might affect the probability that another event E also occurred
  - Reduce the effective sample space from S to F, then measure "size" of E relative to its overlap (if any) in F, rather than relative to S
  - Definition (assuming $P(F) \neq 0$): $$P(E|F) = \frac{P(E \cap F)}{P(F)}$$
- **E and F are *independent* if $P(E \cap F) = P(E)\, P(F)$**
  - Implies $P(E|F) = P(E)$ and $P(F|E) = P(F)$, i.e., knowing that one event occurs tells you nothing about the other
  - If E and F are mutually exclusive, are they independent?

Random Variables

- **One way of quantifying, simplifying events and probabilities**
- **A *random variable* (RV) is a number whose value is determined by the outcome of an experiment**
  - Technically, a function or mapping from the sample space to the real numbers, but can usually define and work with a RV without going all the way back to the sample space
  - Think: RV is a number whose value we don't know for sure but we'll usually know something about what it can be or is likely to be
  - Usually denoted as capital letters: X, Y, W_1, W_2, etc.
- **Probabilistic behavior described by *distribution function***

Discrete vs. Continuous RVs

- **Two basic "flavors" of RVs, used to represent or model different things**
- ***Discrete*** **– can take on only certain separated values**
  - Number of possible values could be finite or infinite
- ***Continuous*** **– can take on any real value in some range**
  - Number of possible values is always infinite
  - Range could be bounded on both sides, just one side, or neither

Discrete Distributions

- **Let X be a discrete RV with possible values (range) x_1, x_2, … (finite or infinite list)**
- ***Probability mass function*** **(PMF)**

$$p(x_i) = P(X = x_i) \quad \text{for } i = 1, 2, \ldots$$

  - The statement "$X = x_i$" is an event that may or may not happen, so it has a probability of happening, as measured by the PMF
  - Can express PMF as numerical list, table, graph, or formula
  - Since X $\sum_{\text{all } i} p(x_i) = 1$ ual to *some* x_i, and since the x_i' s are all distinct,

Discrete Distributions (cont'd.)

- ***Cumulative distribution function*** **(CDF) – probability that the RV will be ≤ a fixed value *x*:**

$$F(x) = P(X \le x) = \sum_{\substack{\text{all } i \text{ such that} \\ x_i \le x}} p(x_i)$$

- **Properties of discrete CDFs**

  $0 \le F(x) \le 1$ for all x

  As $x \to -\infty$, $F(x) \to 0$

  As $x \to +\infty$, $F(x) \to 1$

  $F(x)$ is nondecreasing in x

  These four properties are also true of continuous CDFs

  $F(x)$ is a *step function* continuous from the right with jumps at the x_i's of height equal to the PMF at that x_i

Discrete Distributions (cont'd.)

- **Computing probabilities about a discrete RV – usually use the PMF**
  - Add up $p(x_i)$ for those x_i's satisfying the condition for the event $$P(a < X \le b) = \sum_{\substack{\text{all } i \text{ such that} \\ a < x_i \le b}} p(x_i)$$

- **With discrete RVs, must be careful about weak vs. strong inequalities – endpoints matter!**

Discrete Expected Values

- **Data set has a "center" – the average (mean)** $\bar{X}$
- **RVs have a "center" –** ***expected value***

$$E(X) = \sum_{\text{all } i} x_i p(x_i)$$

  - Also called the *mean* or *expectation* of the RV X
  - Other common notation: μ, μ_X
  - Weighted average of the possible values x_i, with weights being their probability (relative likelihood) of occurring
  - What expectation is *not*: The value of X you "expect" to get

    $E(X)$ might not even be among the possible values $x_1, x_2, \ldots$
  - What expectation *is*:

    Repeat "the experiment" many times, observe many $X_1, X_2, \ldots, X_n$

    $E(X)$ is what $\bar{X}$ converges to (in a certain sense) as $n \to \infty$

Discrete Variances and Standard Deviations

- **Data set has measures of "dispersion" –**
  - Sample variance $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$
  - Sample standard deviation $s = +\sqrt{s^2}$
- **RVs have corresponding measures**

$$Var(X) = \sum_{\text{all } i}(x_i - \mu)^2 p(x_i)$$

  - Other common notation: σ^2, σ_X^2
  - Weighted average of squared deviations of the possible values x_i from the mean
  - Standard deviation of X is $\sigma = \sigma_X = +\sqrt{Var(X)}$
  - Interpretation analogous to that for $E(X)$

Continuous Distributions

- **Now let *X* be a continuous RV**
  - Possibly limited to a range bounded on left or right or both
  - No matter how small the range, the number of possible values for X is always (uncountably) infinite
  - Not sensible to ask about $P(X = x)$ even if x is in the possible range
  - Technically, $P(X = x)$ is always 0
  - Instead, describe behavior of X in terms of its falling *between* two values

Continuous Distributions (cont'd.)

- ***Probability density function*** **(PDF) is a function *f*(*x*) with the following three properties:**

  $f(x) \geq 0$ for all real values x

  The total area under $f(x)$ is 1: $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

  For any fixed a and b with $a \leq b$, the probability that X will fall between a and b is the area under $f(x)$ between a and b: $P(a \leq X \leq b) = \int_a^b f(x)\,dx$

- **Fun facts about PDFs**
  - Observed X's are denser in regions where $f(x)$ is high
  - The height of a density, $f(x)$, is not the probability of anything – it can even be > 1
  - With continuous RVs, you can be sloppy with weak vs. strong inequalities and endpoints

Continuous Distributions (cont' d.)

- ***Cumulative distribution function*** **(CDF) - probability that the RV will be ≤ a fixed value *x*:**

$$F(x) = P(X \le x) = \int_{-\infty}^{x} f(t)\,dt$$

F(x) may or may not have a closed-form formula

- **Properties of continuous CDFs**

$0 \le F(x) \le 1$ for all x

As $x \to -\infty$, $F(x) \to 0$

As $x \to +\infty$, $F(x) \to 1$

$F(x)$ is nondecreasing in x

These four properties are also true of discrete CDFs

$F(x)$ is a continuous function with slope equal to the PDF:

$$f(x) = F'(x)$$

Continuous Expected Values, Variances, and Standard Deviations

- **Expectation or mean of X is**

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

  - Roughly, a weighted "continuous" average of possible values for X
  - Same interpretation as in discrete case: average of a large number (infinite) of observations on the RV X

- **Variance of X is**

$$\sigma^2 = \sigma_X^2 = Var(X) = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x)\, dx$$

- **Standard deviation of X is**

$$\sigma = \sigma_X = +\sqrt{Var(X)}$$

Joint Distributions

- **So far: Looked at only one RV at a time**
- **But they can come up in pairs, triples, …, tuples, forming *jointly distributed* RVs or *random vectors***
  - Input: (T, P, S) = (type of part, priority, service time)
  - Output: $\{W_1, W_2, W_3, \ldots\}$ = output process of times in system of exiting parts
- **One central issue is whether the individual RVs are independent of each other or related**
- **Will take the special case of a pair of RVs (X_1, X_2)**
  - Extends naturally (but messily) to higher dimensions

Joint Distributions (cont' d.)

- *Joint CDF* of (X_1, X_2) is a function of two variables

$$F(x_1, x_2) = P(X_1 \le x_1 \text{ and } X_2 \le x_2) \\ = P(X_1 \le x_1, X_2 \le x_2)$$

  - Same definition for discrete and continuous

- If both RVs are discrete, define the *joint PMF*

$$p(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$$

- If both RVs are continuous, define the *joint PDF* $f(x_1, x_2)$ as a nonnegative function with total volume below it equal to 1, and

$$P(a_1 \le X_1 \le b_1, a_2 \le X_2 \le b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2)\, dx_2\, dx_1$$

- Joint CDF (or PMF or PDF) contains a lot of information – usually won' t have in practice

Marginal Distributions

- **What is the distribution of X_1 alone? Of X_2 alone?**
  - Jointly discrete
    - Marginal PMF of X_1 is $p_{X_1}(x_{1_i}) = P(X_1 = x_{1_i}) = \sum_{\text{all } x_{2_j}} p(x_{1_i}, x_{2_j})$
    - Marginal CDF of X_1 is $F_{X_1}(x) = P(X_1 \le x) = \sum_{\substack{\text{all } i \text{ such} \\ \text{that } x_{1_i} \le x}} p_{X_1}(x_{1_i})$
  - Jointly continuous
    - Marginal PDF of X_1 is $f_{X_1}(x_1) = \int_{-\infty}^{+\infty} f(x_1, x_2)\, dx_2$
    - Marginal CDF of X_1 is $F_{X_1}(x) = P(X_1 \le x) = \int_{-\infty}^{x} f_{X_1}(t)\, dt$
  - Everything above is symmetric for X_2 instead of X_1
- **Knowledge of joint $\Rightarrow$ knowledge of marginals – but not vice versa (unless X_1 and X_2 are independent)**

Covariance Between RVs

- **Measures *linear* relation between X_1 and X_2**
- ***Covariance* between X_1 and X_2 is**

$$Cov(X_1, X_2) = E[(X_1 - E(X_1))(X_2 - E(X_2))]$$

  - If large (resp. small) X_1 tends to go with large (resp. small) X_2, then covariance > 0
  - If large (resp. small) X_1 tends to go with small (resp. large) X_2, then covariance < 0
  - If there is no tendency for X_1 and X_2 to occur jointly in agreement or disagreement over being big or small, then Cov = 0
- **Interpreting value of covariance – difficult since it depends on units of measurement**

Correlation Between RVs

- ***Correlation*** **(coefficient) between X_1 and X_2 is**

$$Cor(X_1, X_2) = \frac{Cov(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}}$$

  - Has same sign as covariance
  - Always between –1 and +1
  - Numerical value does not depend on units of measurement
  - Dimensionless – universal interpretation

Independent RVs

- X_1 and X_2 are *independent* if their joint CDF factors into the product of their marginal CDFs:

$$F(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2) \quad \text{for all } x_1, x_2$$

  - Equivalent to use PMF or PDF instead of CDF
- **Properties of independent RVs:**
  - They have nothing (linearly) to do with each other
  - Independence $\Rightarrow$ uncorrelated
    - But not vice versa, unless the RVs have a joint normal distribution
  - Important in probability – factorization simplifies greatly
  - Tempting just to assume it whether justified or not
- **Independence in simulation**
  - Input: Usually assume separate inputs are indep. – valid?
  - Output: Standard statistics assumes indep. – valid?!?!?!?

Sampling

- ***Statistical analysis*** **– estimate or infer something about a *population* or *process* based on only a *sample* from it**
  - Think of a RV with a distribution governing the population
  - *Random sample* is a set of *independent and identically distributed* (IID) observations $X_1, X_2, \ldots, X_n$ on this RV
  - In simulation, sampling is making some runs of the model and collecting the output data
  - Don't know *parameters* of population (or distribution) and want to estimate them or infer something about them based on the sample

Sampling (cont' d.)

- **Population parameter**
  - Population mean $\mu = E(X)$
  - Population variance σ^2
  - Population proportion
- **Parameter – need to know whole population**
- **Fixed (but unknown)**

- **Sample estimate**
  - Sample mean
  - Sample variance
  - Sample proportion
- ***Sample statistic*** **– can be computed from a sample**
- **Varies from one sample to another – is a RV itself, and has a distribution, called the** ***sampling distribution***

Sampling Distributions

- **Have a statistic, like sample mean or sample variance**
  - Its value will vary from one sample to the next
- **Some sampling-distribution results**
  - Sample mean $\bar{X}$ $\quad E(\bar{X}) = \mu, \quad Var(\bar{X}) = \sigma^2/n$

    If $X \sim N(\mu, \sigma)$ then $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$

    Regardless of distribution of X, $\bar{X} \sim N(\mu, \sigma/\sqrt{n})$ for large n
  - Sample variance s^2
    $E(s^2) = \sigma^2$
  - Sample proportion $\hat{p}$
    $E(\hat{p}) = p$

Point Estimation

- **A sample statistic that estimates (in some sense) a population parameter**
- **Properties**
  - Unbiased: *E*(estimate) = parameter
  - Efficient: *Var*(estimate) is lowest among competing point estimators
  - Consistent: *Var*(estimate) decreases (usually to 0) as the sample size increases

Confidence Intervals

- **A point estimator is just a single number, with some uncertainty or variability associated with it**
- ***Confidence interval*** **quantifies the likely imprecision in a point estimator**
  - An interval that contains (*covers*) the unknown population parameter with specified (high) probability $1 - \alpha$
  - Called a $100\,(1 - \alpha)\%$ confidence interval for the parameter
- **Confidence interval for the population mean μ:**

$$\bar{X} \pm t_{n-1,1-\alpha/2} \frac{s}{\sqrt{n}}$$

$t_{n-1,1-\alpha/2}$ is point below which is area $1 - \alpha/2$ in Student's t distribution with $n - 1$ degrees of freedom

- **CIs for some other parameters – in text**

Confidence Intervals in Simulation

- Run simulations, get results
- View each replication of the simulation as a data point
- Random input $\Rightarrow$ random output
- Form a confidence interval
- Brackets (with probability $1 - \alpha$) the "true" expected output (what you'd get by averaging an infinite number of replications)

Hypothesis Tests

- **Test some assertion about the population or its parameters**
- **Can never determine truth or falsity for sure – only get evidence that points one way or another**
- ***Null hypothesis* (H_0) – what is to be tested**
- ***Alternate hypothesis* (H_1 or H_A) – denial of H_0**

  H_0: $\mu = 6$ vs. H_1: $\mu \neq 6$

  H_0: $\sigma < 10$ vs. H_1: $\sigma \geq 10$

  H_0: $\mu_1 = \mu_2$ vs. H_1: $\mu_1 \neq \mu_2$
- **Develop a decision rule to decide on H_0 or H_1 based on sample data**

Errors in Hypothesis Testing

	H_0 is really true	H_1 is really true
Decide H_0 ("Accept" H_0)	No error Probability $1 - \alpha$ α is chosen (controlled)	Type II error Probability β β is not controlled – affected by α and n
Decide H_1 (Reject H_0)	Type I Error Probability α	No error Probability $1 - \beta$ = power of the test

p-Values for Hypothesis Tests

- **Traditional method is "Accept" or Reject H_0**
- **Alternate method – compute *p-value* of the test**
  - *p*-value = probability of getting a test result more in favor of H_1 than what you got from your sample
  - Small p (like < 0.01) is convincing evidence against H_0
  - Large p (like > 0.20) indicates lack of evidence against H_0
- **Connection to traditional method**
  - If $p < \alpha$, reject H_0
  - If $p \geq \alpha$, do not reject H_0
- ***p*-value quantifies confidence about the decision**

Hypothesis Testing in Simulation

- **Input side**
  - Specify input distributions to drive the simulation
  - Collect real-world data on corresponding processes
  - "Fit" a probability distribution to the observed real-world data
  - Test H_0: the data are well represented by the fitted distribution
- **Output side**
  - Have two or more "competing" designs modeled
  - Test H_0: all designs perform the same on output, or test H_0: one design is better than another
  - Selection of a "best" model scenario