Probability Refresher

Jon Turner
Computer Science & Engineering
Washington University

www.arl.wustl.edu/~jst
Basic Concepts

- Probability theory is study of random experiments
  - example: toss coin 6 times & observe head, tail sequence
- Possible experiment outcomes are sample points, a set of sample points is an event, set of all sample points is sample space
  - HHHHHH is sample point, set of sequences starting with single T is event, and set of all head, tail sequences of length 6 is the sample space ($2^6=64$ points in space)
- Coin toss example involves discrete sample space.
- Can also have continuous sample spaces
  - example: instantaneous voltage measured on an analog communication channel at an arbitrary instant in time
Basic Properties

- For discrete sample space $S$, the *probability distribution* of $S$ is function $\Pr$ that maps sample points to real numbers in $[0,1]$ and satisfies
  \[ \sum_{a \in S} \Pr(a) = 1 \]

- If $A$ and $B$ are events in a discrete sample space,
  \[ \Pr(A) = \sum_{a \in A} \Pr(a) \]
  \[ \Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) \leq \Pr(A) + \Pr(B) \]
  \[ \Pr(A | B) \Pr(B) = \Pr(A \cap B) = \Pr(B | A) \Pr(A) \]

- $A$ and $B$ are *independent* if $\Pr(A | B) = \Pr(A)$
  - in this case, $\Pr(A \cap B) = \Pr(A) \Pr(B)$
Calculating Probabilities Piecewise

- If $H_1, \ldots, H_k$ are events that are mutually exclusive and exhaustive (that is, every sample point is in exactly one of the $H_i$), then

$$\Pr(A) = \sum_{i=1}^{k} \Pr(A \cap H_i) = \sum_{i=1}^{k} \Pr(A|H_i) \Pr(H_i)$$

- Often easier to calculate probabilities of pieces of an event, then sum to get total probability.
Random Variables

- A *random variable* is a function whose domain is the sample space
  - example: the # of heads in a sequence of 6 coin tosses

- For a *discrete random variable* $x_i$,
  - event $x = t$ is set of sample points that $x$ maps to value $t$
  - the probability distribution of $x$ is function $f_x$ that satisfies
    
    \[ f_x(t) = \Pr(x = t) = \sum_{a : x(a) = t} \Pr(a) \]

  - if $x$ is number of heads in a sequence of 6 coin tosses
    \[ f_x(t) = \binom{6}{t} 2^{-6} \]

  - special case of binomial distribution $B(n, p)$ which gives the probability of # of “successes” in sequence of $n$ independent binary trials, where success probability is $p$
Continuous Random Variables

- For a continuous real-valued random variable $x$,
  - the event $x \leq t$ is set of sample points that $x$ maps to values $\leq t$
  - the probability density function for $x$ is the function $f_x$ for which
    \[ \Pr(x \leq t) = F_x(t) = \int_{-\infty}^{t} f_x(z)dz \]
  - $F_x(t)$ is called the cumulative density function of $x$
- The tail probability is
  \[ \Pr(x > t) = 1 - F_x(t) = \int_{t}^{\infty} f_x(z)dz \]
Expected Values

- The expected value (expectation, mean) of a random variable $x$ with distribution (or density function) $f_x$ is denoted $E(x)$

  
  $E(x) = \sum_{t=\infty}^{\infty} tf_x(t)$ (integer valued)

  $E(x) = \int_{\infty}^{\infty} tf_x(t)dt$ (real valued)

- If $x$ is a random variable then $g(x)$ is also, and

  $E(g(x)) = \sum_{t=\infty}^{\infty} g(t)f_x(t)$ or $E(g(x)) = \int_{\infty}^{\infty} g(t)f_x(t)dt$

- If $x_1, \ldots, x_n$ are random variables then

  $E(x_1 + \cdots + x_n) = E(x_1) + \cdots + E(x_n)$
Variance and Standard Deviation

- The variance ($\sigma^2$) of a random variable $x$ is a measure of the probability that $x$ differs significantly from the mean
  \[
  \sigma^2 = E(x^2) - E(x)^2 = E((x - E(x))^2)
  \]
  - the standard deviation of a random variable is $\sigma$
  - for typical distributions, the probability that a r.v. differs from the mean by more than a few $\sigma$ is quite small

- If $z = x + y$, where $x$ and $y$ are independent random variables then $\sigma_z^2 = \sigma_x^2 + \sigma_y^2$
  - so when adding $n$ independent identical random variables, the standard deviation of sum grows as $n^{1/2}$
Common Probability Distributions

- If $x$ is r.v. that counts # of successes in series of $n$ independent binary trials with success probability $p$, then $x$ follows binomial distribution $B(n,p)$,

$$\Pr(x = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \mu = np \quad \sigma^2 = np(1 - p)$$

- If $x$ is r.v. that counts # of trials up to and including first success in series of independent binary trials with success probability $p$, then $x$ follows geometric distribution $G(p)$,

$$\Pr(x = k) = (1 - p)^{k-1} p \quad \mu = 1/p \quad \sigma^2 = (1 - p)/p^2$$
Common Probability Distributions

- *Exponential distribution* can be viewed as a continuous version of the geometric distribution

  \[ f_x(t) = \alpha e^{-\mu t} \quad F_x(t) = 1 - e^{-\mu t} \quad \mu = 1/\alpha \quad \sigma^2 = 1/\alpha^2 \]

  - often used to model packet inter-arrival times and lengths

- *Pareto distribution* is a “heavy-tailed” distribution. For \( t > t_0 \)

  \[ f_x(t) = \alpha t_0^\alpha t^{-(\alpha + 1)} \quad F_x(t) = 1 - (t_0 / t)^\alpha \]

  \[ \mu = \frac{\alpha t_0}{\alpha - 1} \quad \text{for } \alpha > 1 \quad \sigma^2 = \infty \text{ for } 1 < \alpha < 2 \]

  - often used to model burst lengths in the Internet
Memoryless Property

- Consider series of independent binary trials with success probability $p$
  - if $A$ trials pass without success, the distribution of # of trials until next success remains the same, no matter how large $A$ is
    $$\Pr(x \leq A + k | x > A) = \Pr(x \leq k)$$
  - this is the memoryless property

- An exponential r.v. $x$ also has the memoryless property
  $$\Pr(x \leq t + \Delta t | x > t) = \Pr(x \leq \Delta t)$$
  - if packet inter-arrival times are exponentially distributed, time until next packet arrives does not depend on how time since previous packet arrived
Properties of Exponential RVs

- If $x$ and $y$ are independent exponential random variables with parameters $\alpha$ and $\beta$ respectively, then
  \[ \Pr(x \leq y) = \frac{\alpha}{\alpha + \beta} \]

- If $z = \min(x, y)$ is an exponential random variable with parameter $\alpha + \beta$
  - so, if $x$ is the time to next arrival at a queue (mean $1/\lambda$) and $y$ is current outgoing packet finishes (mean $1/\mu$)
  - then $z = \min(x, y)$ is the time until the next arrival or departure and has mean $1/(\lambda + \mu)$
  - and the probability that the arrival comes first is $\lambda / (\lambda + \mu)$
Poisson and Erlang Distributions

- If \( x \) is an exponential random variable with density function \( \alpha e^{-\alpha t} \) that represents the time between successive events, then the number of events \( z \) that occur in a time interval of length \( \tau \) has a Poisson distribution
  \[
  \Pr(z = k) = \frac{(\alpha \tau)^k}{k!} e^{-\alpha \tau} \quad \mu = \alpha \tau \quad \sigma^2 = \alpha \tau
  \]

- If \( x_1, \ldots, x_k \) are identical, independent exponential random variables with density function \( \alpha e^{-\alpha t} \) then \( z = x_1 + \cdots + x_k \) has an Erlang distribution
  \[
  f_z(t) = \frac{\alpha (\alpha t)^{k-1}}{(k-1)!} e^{-\alpha t} \quad F_z(t) = 1 - e^{-\alpha t} \sum_{i=0}^{k-1} \frac{(\alpha t)^i}{i!} \quad \mu = k/\alpha \quad \sigma^2 = k/\alpha^2
  \]
Miscellaneous Useful Facts

■ Stirling’s Approximation
\[
\left(\frac{n}{e}\right)^n \sqrt{2\pi n} \leq n! \leq \left(\frac{n}{e}\right)^n \sqrt{2\pi ne^{1/12n}}
\]

■ Binomial coefficient approximations
\[
\frac{1}{2\sqrt{a}} \leq \frac{1}{2\sqrt{a-b}} \leq \left(\frac{a}{b}\right)^x \leq \left(\frac{a}{b-a}\right)^x \leq \left(\frac{ea}{b}\right)^x
\]

■ From Taylor Series expansion of \(e^x\)
\[
\lim_{x \to 0} e^x = 1 + x
\]

■ Geometric and arithmetic-geometric series
\[
1 + r + r^2 + \cdots + r^{n-1} = \frac{1 - r^n}{1 - r} \quad 1 + r + r^2 + \cdots = \frac{1}{1 - r} \quad (-1 < r < 1)
\]
\[
r + 2r^2 + 3r^3 + \cdots + (n-1)r^{n-1} = \frac{r(1 - nr^{n-1} + (n-1)r^n)}{(1 - r)^2}
\]
\[
r + 2r^2 + 3r^3 + \cdots = \frac{r}{(1 - r)^2} \quad (-1 < r < 1)
\]