StAN Exercise Sheet 1

Richard D. Gill

Mathematical Institute, University of Leiden, Netherlands

http://www.math.leidenuniv.nl/~gill

9 September, 2012

1 Elementary (discrete) probability

1.1 Probability spaces

Construct a (discrete) probability space $(\Omega, \mathcal{F}, \mathrm{P})$ for the experiment "roll a fair die till you get your first six". What is the probability that you have to throw the die for ever (i.e., never ever get a six)?

1.2 Matching problem

A large number n of persons randomly (i.e., *completely* randomly) permute their hats among themselves. Show that the probability that exactly one person gets their own hat is approximately $\exp(-1)$. What can you say about the probability that exactly k persons get their own hat? Do you recognise this probability distribution? Does it make sense?

Hint: Step 1: Use the inclusion-exclusion formula (F & B formula (2.8)) with the event $E_i = \{\text{Person } i \text{ gets their own hat}\}$. What is the event $\bigcup_{i=1}^{n} E_i$?

Step 2: once you have determined the probability that someone gets their own hat, you also know the probability no-one gets their own hat. What about the probability that exactly one person gets their own hat? Let's look at one particular person first. The probability that only Jan gets his own hat is the probability that Jan gets his hat, times the conditional probability that no-one else gets their own hat given that Jan did.

1.3 Repeated conditioning

The definition of $\mathrm{P}(C|A)$, when thought of as a function of all possible events C for a given (fixed) event A, defines a new probability measure on our original sample space. We started with a particular probability measure $\mathrm{P}(\cdot)$ on $(\Omega, \mathcal{F})$ and from that got a new one, $\mathrm{P}(\cdot \mid A)$. One sets all probability outside of A to zero and renormalizes probability inside of A so as to regain total probability mass 1.

Denote by P^A as this conditional probability measure "P given A" got from a given initial probability measure P and a given event A; i.e., for any C, $\mathrm{P}^A(C) := \mathrm{P}(C \mid A)$. Prove that $(\mathrm{P}^B)^A = P^{A \cap B}$; i.e., for any $C \dots$.

This justifies the common notation $\mathrm{P}(C \mid A, B)$ and terminology "probability of C given A and B", which might either be read as the conditional probability got by conditioning first on A, then on B (or first on B, then on A), or as the one-step conditional probability got by conditioning on A and B together.

1.4 Bayes theorem (events)

Suppose A, B and C are any three events. Think of A and B as being two possible causes or antecedents of an observed event C. Prove that

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

In words: posterior odds equals prior odds times likelihood ratio (or Bayes' factor).

Use this result to answer the question: in the standard three doors (Monty Hall) problem, what is the conditional probability that the car is behind door 2, given that the competitor chose door 1 and the host opened door 3 to reveal a goat? What assumptions did you make? Take A and B to be the two events "car is behind door 1" and "car is behind door 2" (two possible "causes"); C is the event "host opens door 3" (the event which we happened to observe). Take it throughout as given that the competitor initially chooses door 1.

2 Random variables, means, distributions

2.1 Variance and mean square

Suppose that X is a random variable with $\mathrm{E}(X^2)$ finite. Defining $\mathrm{Var}(X) := \mathrm{E}((X - \mathrm{E}(X))^2$, show that $\mathrm{Var}(X) = \mathrm{E}(X^2) - \mathrm{E}(X)^2$, using the *linearity* of the expectation operator: for any random variables X and Y with finite means, and for any constants a, b and c, $\mathrm{E}(a + bX + CY) = a + b\mathrm{E}(X) + c\mathrm{E}(Y)$. And of course, a function or transformation of a random variable (for example: "square", "log", "square root") gives us a new random variable.

2.2 Jensen, Markov, Chebyshev

You may assume here that all necessary expectation values exist and are finite. And you may use the fact (given this assumption) that $X \leq Y$ with probability 1 implies that $\mathrm{E}(X) \leq \mathrm{E}(Y)$.

Jensen: Suppose X is a random variable and f is a *convex* function. This last fact implies that through any point on the graph of f, one can draw a straight line which lies everywhere below (or at most only touches) the graph of f. Take $(\mathrm{E}(X), f(\mathrm{E}(X))$ as the point on the graph of f, and let $x \mapsto a + bx$ be the corresponding straight line lying below (or on) the graph of f. So you know that $a + bX \leq f(X)$ with probability one. Use these facts to prove Jensen's inequality: $\mathrm{E}(f(X)) \geq f(\mathrm{E}(X))$.

Markov: Suppose now X is a nonnegative random variable, and $c > 0$ a fixed constant. Show that, when we restrict attention to $x > 0$, the function $x \mapsto cI_{[c,\infty)}(x)$ lies everywhere below the function $x \mapsto x$. Here, $I_B(\cdot)$ stands for the indicator function for the set B, thus equal to $I_B(x) = 1$ for $x \in B$, $= 0$ otherwise. It follows that $cI_{[c,\infty)}(X) \leq X$ with probability one. Use these facts to prove Markov's inequality: $P(X \geq c) \leq \mathrm{E}(X)/c$.

Chebyshev. Suppose $\mathrm{Var}(Y) = \sigma^2 > 0$ and $\mathrm{E}(Y) = \mu$, both finite. Apply Markov's inequality to the random variable $X = ((Y - \mu)/\sigma)^2$ and the constant $c = k^2$ where $k > 0$ is fixed, to prove Chebyshev's inequality: $P(|Y - \mu| \geq k\sigma) \leq k^{-2}$.

2.3 Bell's inequality

Suppose X_1, Y_1, X_2, Y_2 are four random variables which take the values $+1$ and -1 only. Prove that $X_1Y_2 + X_2Y_2 + X_2Y_1 - X_1Y_1 = \pm 2$ with probability 1. Deduce that $\mathrm{E}(X_1Y_2) + \mathrm{E}(X_2Y_2) + \mathrm{E}(X_2Y_1) - E(X_1Y_1)$ cannot exceed 2.

2.4 Infinite means means ...

Suppose U is uniformly distributed on the interval $(0, 1)$, i.e., for any subinterval (a, b), $\mathrm{P}(a < U < b) = b - a$. Define $V = \mathrm{ceiling}(-\log(U)/\log(2))) = -\mathrm{floor}(\log_2(U)))$. Show that $\mathrm{P}(V = k) = 2^{-k}$, $k = 1, 2, \ldots$. Suppose that independently of U, the random variable W takes values ± 1, each with probability $\frac{1}{2}$. Define $X = W2^{V-1}$. Show that X takes the values ± 1, ± 2, ± 4 ... with probabilities 1/4, 1/8, 1/16 ... What is $\mathrm{E}(|X|)$? What is $\mathrm{E}(X)$? Use `R` to study the behaviour of the average of n independent copies of X and of its absolute value $|X|$ as n becomes large. Hint: you might find the functions `runif`, `floor`, `cumsum` useful.

2.5 Bayes theorem (random variables)

Suppose Θ is a random parameter value with probability density $\pi(\theta)$, suppose that X is a random variable such that, conditional on $\Theta = \theta$, X has probability density $p(x|\theta)$. Show that the conditional distribution of Θ given $X = x$ has probability density $\pi(\theta|x) \propto \pi(\theta)p(x|\theta)$ where the constant of proportionality (which can depend on x) is determined by the requirement that $\int_\theta \pi(\theta|x)\mathrm{d}\theta = 1$.

3 Some special distributions

3.1 Binomial distribution

If X has the binomial distribution with parameters n and p, prove that $\mathrm{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$.

Hint: you might find it useful first to compute $\mathrm{E}(X)$ and $\mathrm{E}(X(X-1))$, observing that $\mathrm{E}(X(X-1)) = \mathrm{E}(X^2) - \mathrm{E}(X)$. It is given that $\mathrm{P}(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, \ldots, n$, and that the sum of these probabilities equals 1. This last fact (essentially "Newton's binomial theorem") holds for any natural number n and any probability p.

3.2 Poisson distribution

Suppose X has the Poisson distribution with parameter λ. Show that $\mathrm{E}(X) = \mathrm{Var}(X) = \lambda$.

3.3 Pareto distribution – Salpeter function

Suppose X has the probability density $c(\alpha, x_0)x^{-\alpha}$ on $x > x_0$; $x_0 > 0$ and $\alpha > 0$ given parameters. What is $c(\alpha, x_0)$? What are the (cumulative) distribution function F of X and the quantile function Q of X? What are mean and variance of X? What is the distribution of $\log(X/x_0)$? (Hint: answer all these questions in the case $x_0 = 1$ first, and derive the general case from this.)

3.4 Gamma distribution – Schechter function

Suppose X has the probability density $c(\alpha, \lambda)x^{-\alpha-1}e^{-\lambda x}$ on $x > 0$. Show that $c(\alpha, \lambda) = \lambda^\alpha / \Gamma(\alpha)$ where $\Gamma(\alpha) := \int_{y=0}^{\infty} y^{\alpha-1} e^{-y}$ for $\alpha > 0$ is the gamma function; note that for positive integers n, $\Gamma(n) = (n-1)!$ (proof: integration by parts and mathematical induction!). Compute $\mathrm{E}(X)$, $\mathrm{E}(X^2)$, $\mathrm{Var}(X)$. Hint: do all this in the case $\lambda = 1$ first! Then show that if Y has the gamma distribution with parameters α and 1, then Y/λ has the gamma distribution with paramours α and λ.

Now suppose that $x_0 > 0$ is fixed, and let U be a uniformly $(0, 1)$ distributed random variable, independent of X. Consider the conditional probability distribution of X given $X > y_0$ and $U < y_0/X$ where $y_0 > 0$ is fixed. Let Y be a random variable with this distribution distribution. Define $\beta = \alpha - 2$. Show that Y has probability density $c(\beta, \lambda, y_0)y^{-\beta}e^{-\lambda y}$ on $y > y_0$. Express the normalisation constant c in terms of gamma probability density functions and distribution functions.