Prof. Donna Ankerst, Katharina Selig

May 6th & 7th 2019

Problem 1

A discrete random variable X has the following probability mass function

x	0	1	2	3	4
p(x)	0.1	0.2	0.2	0.2	0.3

(a) Write out the inverse transform algorithm for this problem by hand in terms of the specific numbers.

`In R:`

(b) Use the inverse transform method to generate a random sample of size 1000 from the distribution of X.

(c) Construct a relative frequency table and compare the empirical distribution of your sample with the theoretical probabilities.

(d) Generate another sample of size 1000 from the distribution of X by using the `sample` function and plot a histogram of this sample.

Problem 2

The Rayleigh(σ) density is given by

$$f(x) = \frac{x}{\sigma^2} e^{-x^2/2\sigma^2}, \quad x \geq 0, \sigma > 0$$

(a) Develop and write by hand an algorithm to generate random variables from a Rayleigh(σ) distribution.

(b) Write a `R` function to generate a random variable using the algorithm in (a).

`In R:`

(c) Generate 1000 Rayleigh(σ) samples for $\sigma = 0.5, 1, 2, 4$ and plot the density estimate using the `R` function `geom_density` from `ggplot2`. How does the parameter σ influence the shape of the histogram?

Problem 3

The density of the Beta(α, β) distribution is given by

$$f(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \leq x \leq 1, \alpha > 0, \beta > 0$$

with the Gamma function $\Gamma(n) = (n-1)!$ for positive integers n.

(a) Develop and write out an acceptance rejection algorithm specifically for a Beta(2,5)-distribution.

(b) Write `R` code for the algorithm in (a) for a given sample size of `n`.

(c) What is the anticipated acceptance probability for your algorithm as well as the average number of iterations to obtain 1000 random samples?

`In R:`

(d) Generate a sample of size 1000 and plot a histogram overlaying the theoretical distribution. State the number of iterations required and compare to (c).

Problem 4

A rainfall company would like to design an experiment to measure the daily rainfall over a range of regions in the Amazon, where the rainfall measurements in cm are expected to follow an $\text{Exp}(\frac{1}{2})$ distribution. They want to know the number of regions (n) that they need to measure to estimate both the mean and 0.95-quantile with standard errors (ste) within 20% of the estimated mean and the 0.95-quantile, respectively. Recall that a random variable $X \geq 0$ that follows a $\text{Exp}(\lambda)$ distribution with density $f(x, \lambda) = \lambda \exp(-\lambda x)$ has mean $\frac{1}{\lambda}$ and variance $\frac{1}{\lambda^2}$.

(a) Compute by hand the necessary sample size for the mean.

(b) Write the formula for the sample size for the 0.95-quantile.

`In R:`

(c) Calculate (b) and compare to (a).

(d) Draw two times a sample of the $\text{Exp}(\frac{1}{2})$ distribution, one with the sample size calculated in (a) and the other with the sample size as calculated in (c). Plot a histogram for each sample and overlay the theoretical distribution. Explain the difference between the sample size in (a) and (b) using the graphs.

Problem 5

ProSieben would like to design a survey to estimate the number of households that watch Germany's Next Top Model (GNTM). Specifically, they would poll n households and assume the number X if those that watch the show is $\text{Binomial}(n, \pi)$ distributed, where π is the population proportion that watch the show. The estimated proportion $p = \frac{X}{n}$ has $ste(p) = \sqrt{\frac{\pi(1-\pi)}{n}}$ and an approximate $100(1-\alpha)\%$ CI for π would be given by $p \pm z_{1-\frac{\alpha}{2}} ste(p)$, where $z_{1-\frac{\alpha}{2}}$ is the $1 - \frac{\alpha}{2}$ quantile of the standard distribution.

(a) Derive the sample size (n) needed to control the half-width of the CI by bound B when the true proportion is π.

`In R:`

(b) Plot the sample size as a function of π for bound $B = 0.03$ and a 95% CI. Determine the π corresponding to the maximum sample size.

(c) Suppose ProSieben says that based on past years, the anticipated proportion is $\pi = 0.15$. Compare this sample size to the maximum from (b), which should be used as a conservative estimate when the company has no idea. Calculate the bound B that will be obtained if ProSieben uses the max sample size from (b) but $\pi = 0.15$ was later observed and compare those to their requirement of $B = 0.03$.

(d) Overlay on plot (b) the sample size curve assuming a 99% CI instead of 95% and compare.

Problem 6

(a) Calculate the mean and variance of a $\text{Ber}(p)$ distribution, showing all steps.

(b) Suppose X_1 and X_2 are independent continuous random variables with densities f_{X_1} and f_{X_2}, respectively. Show by integrals that the mean of their sum is the sum of their means and variance of their sum is the sum of their variances. Suppose they are not independent. Do any of these two properties hold?

(c) Suppose $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$. Use the results in (b) to obtain the mean and variance of $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.