Probability Distributions
EBP038A05, 2021-2022.2A
Mock exam, April 4 2022
Student id:
Student name:

dr. N.D. van Foreest c.s. FEB, University of Groningen

Read-me

1. The exam is closed book.
2. The exam duration is 2 hours
3. Each exam question contains one or more sub-questions. For each sub-question you can earn 0 (no answer, or completely wrong), 1, 2 or 3 (perfect) points.
4. Write your answer below the question. (If you need scrap paper, just use the backside of the pages.)
5. For numerical answers, rounding to 3 significant digits is sufficient. However, stating answers like 11/19 is preferred. Actually, you don't need a calculator at all.

For this mock exam, we don't include solutions here. You can find problems in the book, and look up the solutions in the study guide. For the code questions, use the assignments.

1 Question

Customers arrive at a shop in accordance to a Poisson process with rate λ per hour. Each customer buys an item with probability p, independent of other customers. The purchase price of an item has mean μ and variance σ^2.

Ex 1.1. What is the amount spent by a random customer (including the customers that don't buy)?

Ex 1.2. What is the mean and the variance of the revenue of the shop received during a morning of 4 hours long?

2 Question

Let U and V be iid and geometrically distributed with success probability p. Take $N = U + V$.

Ex 2.1. Find $\mathsf{E}[N]$.

Ex 2.2. Write down the joint distribution of U and V.

Ex 2.3. Find an expression for $\mathsf{P}\{U = i | N = n\}$.

3 Question

Let $\{X_i\}$ be a sequence of iid rvs such that $X_i \sim \text{Unif}(0,1)$. Suppose that $Y_i = -\lambda^{-1}\log(1/X_i)$ for all i.

Ex 3.1. How is $Y_1 \times \cdots \times Y_n$ distributed, and what are the parameters?

4 Question

This simulation exercise is based on BH.9.25. For the the exam we will copy the text of the question, but for the mock exam we expect you to look it up in BH.

Python Code

```python
import numpy as np
from scipy.stats import bernoulli

np.random.seed(3)

n = 5
num = 10

p = 0.5
S = bernoulli(p).rvs([num, n]) * 2 - 1 # this 1

x = np.zeros([num, n])
x[:, 0] = 100
f = 0.25

for i in range(1, n):
    x[:, i] = x[:, i - 1] +  (1 - f * S[:, i]) # this 2

print(x.mean(axis=0), x.std(axis=0))
```

R Code

```r
set.seed(3)

n = 5
num = 10

p = 0.5
S = matrix(0, num, n)
for(i in 1:num){
  S[i,1:n] = rbinom(n, 1, p) * 2 -1 #this 1
}

x = matrix(0, num, n)
x[,1] = 100
f = 0.25

for(i in 2:n){
```

```
  x[,i] = x[,i-1] + (1 - f * S[,i])  # this 2
}

print(colMeans(x))
print(apply(x, MARGIN = 2, sd))
```

Ex 4.1.

1. Line ‘this 1’: explain why we multiply by 2 and subtract 1.
2. Line ‘this 2’: this line contains errors. How should it be repaired so the code correctly computes the answer for the question.

5 Question

Let X and Y be iid continuous rvs. We use the code below

1. to get an indication of whether $\mathsf{P}\{X > Y+3\} \geq \mathsf{P}\{Y > X+3\}$;
2. to estimate $\mathsf{E}[XY]$.

Python Code

```python
import numpy as np
from scipy.stats import expon

np.random.seed(3)

X = expon(2)
Y = expon(3)

n_sample = 1000
X_sample = X.rvs(n_sample)
Y_sample = Y.rvs(n_sample)
p1 = sum(X.rvs(n_sample) - Y.rvs(n_sample) + 3) / n_sample
p2 = sum(Y_sample  X_sample > 3) / n_sample
print(p1-p2)

EXY = X_sample @ Y_sample / n_sample # this
```

R Code

```r
library(cubature)
set.seed(3)

n_sample = 1000
X_sample = rexp(n_sample, 1) + 2
Y_sample = rexp(n_sample, 1) + 3
p1 = sum((rexp(n_sample, 1) + 2) - (rexp(n_sample, 1) + 3) + 3) / n_sample
p2 = sum(Y_sample  X_sample > 3) / n_sample
print(p1 - p2)

EXY = X_sample %*% Y_sample / n_sample #this
```

Ex 5.1.

1. This code does not print $\mathsf{P}\{X > Y+3\} - \mathsf{P}\{Y > X+3\}$. What is wrong? Explain how to repair it (or provide the correct code).
2. In the ‘this’ line we use simulation to estimate $\mathsf{E}[XY]$. What is the problem of using a numerical integrator for this task?

List of distributions

For the following distributions, you have to know by heart the form and the parameters, and either learn (or be able to derive at the exam) the mean and variance: Bernoulli, Binomial, First success, Geometric, Poisson, Uniform (discrete and continuous) and Exponential.

For the hypergeometric distribution you have to know the pmf and the parameters, but not the mean and variance. If necessary, we will provide the mean and variance at the exam.

We will not ask any question that involves calculus (e.g., integration) with (the cdf or pmf of): Negative hyper geometric, Weibull, Log normal, Chi-square or Student-t.

name	pmf	μ	σ^2
NBin	$\binom{r+k-1}{r-1} p^r q^k$	rq/p	rq/p^2
Normal	$\frac{e^{-(x-\mu)^2/2\sigma^2}}{\sqrt{2\pi\sigma^2}}$	μ	σ^2
Gamma	$\frac{(\lambda x)^a e^{-\lambda x}}{x\Gamma(a)}$	a/λ	a/λ^2
Beta	$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}$	$\mu = \frac{a}{a+b}$	$\frac{\mu(1-\mu)}{a+b+1}$,

where $\Gamma(a) = (a-1)!$ if a is a positive integer.

LAST PAGE OF THE EXAM