

# Statistical Methods - I partial test

November 8, 2023 - 1 hour

The text must be submitted with name and surname to allow evaluation.

|                  |            |   |
|------------------|------------|---|
| Name and Surname | Student ID | 1 |
|------------------|------------|---|

**1** A discrete random variable  $X$  ranging from 0 to 3 has the following probabilities  $P(X = 0) = 0.027$ ,  $P(X = 1) = 0.189$ ,  $P(X = 2) = 0.441$ . Find the mean and the variance of such variable.

- ☐  $E(X) = 2.1$ ;  $V(X) = 5.04$
- ☐  $E(X) = 2.1$ ;  $V(X) = 0.63$
- ☐  $E(X) = 3.1$ ;  $V(X) = 0.63$
- ☐  $E(X) = 3.1$ ;  $V(X) = 0.79$

**2** Which sentence is true for a random variable  $X \sim \text{Binom}(30, 0.9)$

- ☐ The distribution is asymmetric and the variance is the highest possible
- ☐ The distribution is symmetric and the variance is the highest possible
- ☐ The distribution is asymmetric and the variance is not the highest possible
- ☐ The distribution is symmetric and the variance is not the highest possible

**3** A 95% confidence interval obtained from a random sample of 500 people has

- ☐ We do not have enough information to make comparisons with other situations
- ☐ A better chance of containing the population mean than a 95% CI obtained from a random sample of 1000 people
- ☐ Less of a chance of containing the population mean than a 95% CI obtained from a random sample of 1000 people
- ☐ The same chance of containing the population mean than a 95% CI obtained from a random sample of 1000 people

**4** Given a bivariate Normal random variable  $(X, Y)$  with mean vector  $\boldsymbol{\mu} = (1, -1)$ , variances  $\sigma_x^2 = 2$  and  $\sigma_y^2 = 0.5$ , and correlation  $\rho = 0.75$ , find the distribution of the random variable  $T = 2X - 4Y$ .

- ☐  $T \sim N(6, 16)$
- ☐  $T \sim N(6, 4)$
- ☐  $T \sim N(-2, 1.5)$
- ☐  $T \sim N(-2, 12)$

**5** Given a random sample from a normal population of size  $n = 100$ , mean  $\bar{x} = 23$ , and variance  $S^2 = 144$ , find the confidence interval for the mean for a level  $\alpha = 0.01$ .

- ☐ (19.909, 26.091)
- ☐ (14.16, 31.84)
- ☐ (19.291, 26.709)
- ☐ (15.273, 30.727)

**6** An estimator,  $\hat{\theta}_n$ , for a parameter  $\theta$ , is said to be asymptotically efficient if, as  $n \rightarrow \infty$

- ☐  $E[\hat{\theta}_n - \theta] \rightarrow 0$  and  $E[(\hat{\theta}_n - E(\hat{\theta}_n))^2] \rightarrow \mathcal{I}(\theta)^{-1}$
- ☐  $E[\hat{\theta}_n - E(\hat{\theta}_n)] \rightarrow 0$
- ☐  $E[\hat{\theta}_n - \theta] \rightarrow 0$
- ☐  $E[(\hat{\theta}_n - \theta)^2] \rightarrow 0$

7 Figure 1 represents the empirical cumulative distribution function of a sample generated from which of the following distributions.

- ☐ Standard Normal
- ☐ Gamma
- ☐ Beta
- ☐ Binomial

8 Code 2 shows the Italian population distribution for the blood types BTpop and the empirical conditional distributions from three samples of size 2500 each from the three Italian macro regions (South, Central, North), BTobs. The following code is meant to establish whether:

```
chisq.test(BTobs[1,], p = BTpop)
```

- ☐ A specific blood type is more prevalent in the Southern regions
- ☐ The observed frequency of blood types of the Southern people differs from the Italian distribution
- ☐ The blood type of the Southern people is uniformly distributed
- ☐ The blood types are homogeneously distributed among the regions

9 Code 1 is meant to assess via Monte Carlo simulation the exact or approximate distribution for the sum of i.i.d. random variables. Let a and b be the location and scale parameters of the random variables, which R code draws the density curve overlapping the histogram?

- ☐ `curve(dnorm(x, n * b, n * b), add = TRUE)`
- ☐ `curve(dnorm(x, n * a, sqrt(n * b^2)), add = TRUE)`
- ☐ `curve(dgamma(x, shape = n, scale = b), add = TRUE)`
- ☐ `curve(dnorm(x, n * a, sqrt(n * b)), add = TRUE)`

10 In the following R code x is the vector of observations and theta is a scalar denoting the parameter of interest. What does the code implement?

```
llik <- function(theta, x) -length(x) * log(theta)/2 - sum(x^2)/(2*theta)
```

- ☐ The log-likelihood function for  $\sigma^2$  for a random sample from  $X \sim N(\mu, \sigma^2)$ , with  $\mu = 0$
- ☐ The log-likelihood function for  $\pi$  for a random sample from  $X \sim Binom(1, \pi)$
- ☐ The log-likelihood function for  $\lambda$  for a random sample from  $X \sim Pois(\lambda)$
- ☐ The log-likelihood function for  $\alpha$  for a random sample from  $X \sim Gamma(\alpha, \beta)$ , with  $\beta$  known

#### Code 1 - MC simulation

```
n <- 100
R <- 10000
a <- 3
b <- 2
sample <- rep(NA, R)
for(i in 1 : R){
  u <- runif(n)
  sample[i] <- sum(a + b * qnorm(u))
}
hist(sample, prob = TRUE)
```

#### Code 2 - Blood Type

```
> BTpop
  A   B  AB   O
0.36 0.17 0.07 0.40

> BTobs
      A   B  AB   O
South  475 700 300 1025
Central 1125 300 75 1000
North  1100 275 125 1000
```

Figure 1

