5.

Setup: All problems on this page will follow the definitions here:

Let X, Y be two Bernoulli random variables and let

$$\begin{aligned} p &= P(X=1) \qquad (\text{the probability that } X=1) \\ q &= P(Y=1) \qquad (\text{the probability that } Y=1) \\ r &= P(X=1, Y=1) \qquad (\text{ the probability that both } X=1 \text{ and } Y=1)\,. \end{aligned}$$

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sample of n i.i.d. copies of (X, Y). Based on this sample, we want to test whether X and Y are independent, i.e., whether $r = pq$.

Independence of X and Y, Part 1

0.5/0.5 points (graded)

If X and Y are independent, then what is r in terms of p and q?

$r =$ p*q

$p \cdot q$

✔ **Answer:** p*q

Submit You have used 1 of 3 attempts

Generating Speech Output

Answers are displayed within the problem

Independence of X and Y, Part 2

2.0/2.0 points (graded)

With no assumptions of X and Y except for the definitions above (i.e., do not assume X and Y are independent), find the following probabilities

0. $P(X=1, Y=0)$ in terms of p and r.

$P(X=1, Y=0) =$ p-r ✔ **Answer:** p-r

$p-r$

0. $P(X=0, Y=1)$ in terms of q and r.

$P(X=0, Y=1) =$ q-r ✔ **Answer:** q-r

$q-r$

0. $P(X=0, Y=0)$ in terms of p, q and r.

$P(X=0, Y=0) =$ 1+r-p-q ✔ **Answer:** 1-p-q+r

$1+r-p-q$

If $r = pq$, then which of the following is true?

- (selected) X and Y are independent.
- X and Y are not independent.

✔

Generating Speech Output

Solution:

By definition of joint probabilities:

$$P(X=1,Y=0) \;=\; P(X=1)-P(X=1,Y=1) \;=\; p-r.$$

Similarly,

$$\begin{aligned}P(X=0,Y=1) &= P(Y=1)-P(X=1,Y=1) \;=\; q-r\\ P(X=1,Y=1) &= 1-P(X=1,Y=0)-P(X=0,Y=1)-P(X=0,Y=0)\\ &= 1-(p-r)-(q-r)-r \;=\; 1-p-q+r.\end{aligned}$$

Hence, if given $r=pq,$ then

$$\begin{aligned}P(X=0,Y=0) &= r=pq \;=\; P(X=0)\,P(Y=0)\\ P(X=1,Y=0) &= p-r \;=\; p(1-q) \;=\; P(X=1)\,P(Y=0)\\ P(X=0,Y=1) &= q-r \;=\; q(1-p) \;=\; P(X=0)\,P(Y=1)\\ P(X=1,Y=1) &= 1-p-q+r \;=\; (1-p)(1-q) \;=\; P(X=1)\,P(Y=1)\,.\end{aligned}$$

Therefore, the single equation $r=pq$ guarantees the independence of X and Y.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Estimators of p, q, and r

2.0/2.0 points (graded)

Note: In all problems below, do not assume $r=pq$.

Define

$$\hat{p} \;=\; \frac{1}{n}\sum_{i=1}^{n} X_i$$

Generating Speech Output

$$\hat{q} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

$$\hat{r} = \frac{1}{n}\sum_{i=1}^{n} X_i Y_i.$$

Are these consistent estimators of p, q and r respectively? Find the limits $\lim_{n\to\infty} \hat{p}$, $\lim_{n\to\infty} \hat{q}$, and $\lim_{n\to\infty} \hat{r}$.

$\lim_{n\to\infty} \hat{p} =$ p ✔ **Answer:** p

p

$\lim_{n\to\infty} \hat{q} =$ q ✔ **Answer:** q

q

$\lim_{n\to\infty} \hat{r} =$ r ✔ **Answer:** r

r

Is the vector $\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{r} \end{pmatrix}$ asymptotically normal?

- (●) Yes
- ( ) No
- ( ) Not enough information to decide

✔

Solution:

Generating Speech Output

- By the (weak) law of large numbers:

$$\begin{aligned}
\hat{p} &= \overline{X_n} \xrightarrow[n\to\infty]{(P)} p \\
\hat{q} &= \overline{Y_n} \xrightarrow[n\to\infty]{(P)} q \\
\hat{r} &= \frac{1}{n}\sum_{i=1}^{n} X_i Y_i \xrightarrow[n\to\infty]{(P)} P(XY=1) = P(X=1, Y=1) = r
\end{aligned}$$

where in the last equation the product XY remaind a Bernoulli random variable.

- The vector

$$\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{r} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \begin{pmatrix} X_i \\ Y_i \\ X_i Y_i \end{pmatrix}$$

is an average, hence we can conclude by the CLT that it is asymptotically normal, i.e. converges in distribution to a Multivariate Gaussian.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Asymptotic Covariance Matrix

2.5/2.5 points (graded)

As above, consider the vector $\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{r} \end{pmatrix}$.

Find the asymptotic covariance matrix Σ of the vector $\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{r} \end{pmatrix}$ in terms of $p,\ q,$ and r. That is, find the covariance matrix of $\sqrt{n}\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{r} \end{pmatrix}$ as $n \to \infty$. (Do not assume $r = pq$.)

Generating Speech Output

(Enter your answer as a matrix, e.g. by typing **[[1,2],[5*x,y-1]]** for the matrix $\begin{pmatrix} 1 & 2 \\ 5x & y-1 \end{pmatrix}$. Note the **square brackets**, and **commas as separaters**.)

$\Sigma =$ [[p*(1-p),r-p*q,r-p*r],[r-p*q,q*(1-q),r-q*r],[r-p*r,r-q*r,r*(1-r)]]

Answer: [[p*(1-p),r-p*q,r*(1-p)],[r-p*q,q*(1-q),r*(1-q)],[r*(1-p),r*(1-q),r*(1-r)]]

Solution:

By the CLT,

$$\sqrt{n}\left[\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{r} \end{pmatrix} - \begin{pmatrix} p \\ q \\ r \end{pmatrix}\right] \xrightarrow[n\to\infty]{(d)} \mathcal{N}\left(\mathbf{0}, \Sigma\right)$$

where Σ is the covariance matrix of $\begin{pmatrix} X \\ Y \\ XY \end{pmatrix}$:

$$\Sigma = \begin{pmatrix} X \\ Y \\ XY \end{pmatrix} = \begin{pmatrix} \mathsf{Var}\left(X\right) & \mathsf{Cov}\left(X,Y\right) & \mathsf{Cov}\left(X,XY\right) \\ \mathsf{Cov}\left(Y,X\right) & \mathsf{Var}\left(Y\right) & \mathsf{Cov}\left(Y,XY\right) \\ \mathsf{Cov}\left(XY,X\right) & \mathsf{Cov}\left(XY,Y\right) & \mathsf{Var}\left(XY\right) \end{pmatrix}.$$

Since X, Y, XY are Bernoulli with paremeter p, q, r respectively , the diagonal terms (i.e. their respective variances) are $p\left(1-p\right)$, $q\left(1-q\right)$, $r\left(1-r\right)$. The covariances are:

$$\begin{aligned} \mathsf{Cov}\left(X,Y\right) &= P\left(X=1,Y=1\right) - P\left(X=1\right)P\left(Y=1\right) = r - pq \\ \mathsf{Cov}\left(X,XY\right) &= P\left(X=1,XY=1\right) - P\left(X=1\right)P\left(XY=1\right) = P\left(X=1,Y=1\right) - P\left(X=1\right)P\left(XY=1\right) = r - pr \\ \mathsf{Cov}\left(Y,XY\right) &= P\left(Y=1,XY=1\right) - P\left(X=1\right)P\left(XY=1\right) = P\left(X=1,Y=1\right) - P\left(Y=1\right)P\left(XY=1\right)\ r - qr. \end{aligned}$$

Submit You have used 1 of 3 attempts

Generating Speech Output Answers are displayed within the problem

Asymptotic Variance

1.0/1.0 point (graded)

Use the Delta method to find the asymptotic variance V of $\hat{r} - \hat{p}\hat{q}$ in terms of $p,\ q$ and r. Specify the vector w such that $V = w^T \Sigma w,$ where Σ the asymptotic covariance matrix of $\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{r} \end{pmatrix}$ from the problem above,

$$\sqrt{n}\left(\hat{r} - \hat{p}\hat{q} - (r - pq)\right) \xrightarrow[n\to\infty]{(d)} \mathcal{N}\left(0, V\right).$$

(Enter your answer as a vector, e.g. type **[5*x,y-1,3]** for the vector $\begin{pmatrix} 5x & y-1 & 3 \end{pmatrix}$ or its transpose. Note the **square brackets**, and **commas as separaters**.)

$w =$ [-q, -p, 1] ✔ **Answer:** [-q,-p,1]

Solution:

Recall that by the CLT,

$$\sqrt{n}\left[\begin{pmatrix} \hat{p} \\ \hat{q} \\ \hat{r} \end{pmatrix} - \begin{pmatrix} p \\ q \\ r \end{pmatrix}\right] \xrightarrow[n\to\infty]{(d)} \mathcal{N}\left(\mathbf{0}, \Sigma\right).$$

where Σ is the covariance matrix of $\begin{pmatrix} X \\ Y \\ XY \end{pmatrix}$. For the delta method, define

$$\begin{aligned} g : \mathbb{R}^3 &\rightarrow \mathbb{R} \\ \begin{pmatrix} p \\ q \\ r \end{pmatrix} &\mapsto r - pq \end{aligned}$$

Generating Speech Output adient is

$$\nabla g = \begin{pmatrix} -q \\ -p \\ 1 \end{pmatrix}.$$

Then the delta method gives

$$\sqrt{n}\left(g\begin{pmatrix}\hat{p}\\ \hat{q}\\ \hat{r}\end{pmatrix} - g\begin{pmatrix}p\\ q\\ r\end{pmatrix}\right) = \sqrt{n}\left[(\hat{r}-\hat{p}\hat{q})-(r-pq)\right] \xrightarrow[n\to\infty]{(d)} \mathcal{N}\left(\mathbf{0}, \left[\nabla g\begin{pmatrix}p\\ q\\ r\end{pmatrix}\right]^T \Sigma\left(\nabla g\begin{pmatrix}p\\ q\\ r\end{pmatrix}\right)\right] = \mathcal{N}\left(\mathbf{0}, \left(\begin{pmatrix}-q & -p & 1\end{pmatrix}\Sigma\begin{pmatrix}-q\\ -p\\ 1\end{pmatrix}\right)\right).$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Asymptotic Variance Under the Null

1.5/1.5 points (graded)

Consider the following hypotheses:

$$H_0 : X \text{ and } Y \text{ are independent} \quad \text{vs.} \quad H_1 : X \text{ and } Y \text{ are not independent.}$$

Assuming that H_0 is true, find the asymptotic variance V of $\hat{r} - \hat{p}\hat{q}$ in terms of p and q only.

Hint: Rephrase the hypotheses in terms of the parameters $p,\ q,$ and r.

$V =$ p*(1-p)*q*(1-q) ✔ **Answer:** p*q*(1-p)*(1-q)

$p \cdot (1-p) \cdot q \cdot (1-q)$

Propose a consistent estimator $\hat{V}$ of V. (There is no answer box for this question to avoid double jeopardy.)

Solution:

Generating Speech Output

We have shown earlier that X and Y are independent if and only if $r = pq$, hence the null hypothesis can be restated as $H_0 : r - pq = 0$.
Recall the asymptotic variance of $\hat{r} - \hat{p}\hat{q}$ is

$$\begin{pmatrix} -q & -p & 1 \end{pmatrix} \begin{pmatrix} p(1-p) & r-pq & r(1-p) \\ r-pq & q(1-q) & r(1-q) \\ r(1-p) & r(1-q) & r(1-r) \end{pmatrix} \begin{pmatrix} -q \\ -p \\ 1 \end{pmatrix}.$$

Under $H_0 : r - pq = 0$, this reduces to

$$V = \begin{pmatrix} -q & -p & 1 \end{pmatrix} \begin{pmatrix} p(1-p) & 0 & r(1-p) \\ 0 & q(1-q) & r(1-q) \\ r(1-p) & r(1-q) & r(1-r) \end{pmatrix} \begin{pmatrix} -q \\ -p \\ 1 \end{pmatrix} = pq(1-p)(1-q)$$

(where we have used software to compute the matrix product).
Hence, a consistent estimator $\hat{V}$ of V is $\hat{p}\hat{q}\,(1-\hat{p})(1-\hat{q})$.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Test for Independence

2.0/2.0 points (graded)
As above, consider the following hypotheses:

$$H_0 : X \text{ and } Y \text{ are independent} \quad \text{vs.} \quad H_1 : X \text{ and } Y \text{ are not independent.}$$

Propose a test Ψ for the hypotheses above with asymptotic level α, for any $\alpha \in (0,1)$. Let $\Psi = \mathbf{1}\left(|T_n| > q_{\alpha/2}\right)$, where $q_{\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. Specify the test statistic T_n below. Your test statistic should be in terms of n, $\hat{r}$, $\hat{p}$, $\hat{q}$, and if desired, also $\hat{V}$, the estimator of the asymptotic variance of $\hat{r} - \hat{p}\hat{q}$ assuming H_0 is true.

(Type **hatr** for $\hat{r}$,**hatp** for $\hat{p}$,**hatq** for $\hat{q}$, and **hatV** for $\hat{V}$, the estimator of the asymptotic variance of $\hat{r} - \hat{p}\hat{q}$ under the null.)

Generating Speech Output

$T_n =$ sqrt(n)*(hatr-hatp*hatq)/sqrt(hatV) **Answer:** sqrt(n)*(hatr-hatp*hatq)/sqrt(hatV)

Solution:

A two-sided test $\Psi = \mathbf{1}\left(|T_n| > q_{\alpha/2}\right)$, where $q_{\alpha/2}$ would have the test statistic

$$T_n \;=\; \sqrt{n}\frac{\hat{r} - \hat{p}\,\hat{q}}{\sqrt{\hat{V}}}.$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Happiness and Being in a Relationship

1.5/1.5 points (graded)

We would like to know whether the facts of being happy and being in a relationship are independent of each other. In a given population, 1000 people (aged at least 21 years old) are sampled randomly with replacement and asked two questions: "Do you consider yourself as happy ?" and "Are you involved in a relationship ?". The answers are summarized in the following table:

	Happy	Not happy
In a relationship	205	301
Not in a relationship	179	315

Denote by p, q and r the true proportions of people that are happy, in a relationship, and both happy and in a relationship, respectively.

Compute the values of $\hat{p}$, $\hat{q}$ and $\hat{r}$.

$\hat{p} =$ 0.384 **Answer:** 0.384

Generating Speech Output

$\hat{q} =$ 0.506 ✔ **Answer:** 0.506

$\hat{r} =$ 0.205 ✔ **Answer:** 0.205

Solution:

$$\begin{aligned} \hat{p} &= \frac{205+179}{1000} = 0.384 \\ \hat{q} &= \frac{205+301}{1000} = 0.506 \\ \hat{r} &= \frac{205}{1000} = 0.205 \end{aligned}$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

P-value

1.5/1.5 points (graded)

Does the test that you defined in the problem above invalidate the independence of being happy and being in a relationship at asymptotic level 5% ?

- ( ) Invalidates the independence of being happy and being in a relationship
- (•) Does not invalidate the independence of being happy and being in a relationship

✔

What is the p-value of that test ? (Enter an answer accurate to at least 2 decimal places.)

p-value: 0.1642259 ✔ **Answer:** 0.164

Generating Speech Output

Solution:

Given the data, the test statistic T_n is evaluated to be:

$$T_n = \sqrt(n)\frac{\hat{r} - \hat{p}\,\hat{q}}{\sqrt{\hat{p}\,\hat{q}\,(1-\hat{p})\,(1-\hat{q})}} = 1.3910 < q_{0.025}.$$

This gives p-value 0.164.
Hence, we fail to reject H_0, i.e. we cannot invalidate the independence of happiness and being in a relationship.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Error and Bug Reports/Technical Issues

Hide Discussion

Topic: Final exam:Final Exam / 5.

Add a Post

Show all posts ▼ by recent activity ▼

[Staff] partial credit to Asymptotic Covariance Matrix please
The question worths 2.5 points. I think we deserve partial credit for this question 2

[Staff] double jeopardy
The answers to the Asymptotic Variance Under the Null and the P-Value questions depend on the Asymptotic Covariance Matrix question. Would it be possible to grade th... 1

? [STAFF] Absolute value in test for independence 1

My Answers to this question(4) 21

Resolution of the Final Exam in Julia 1

[Staff] Thank you!
Thank you ! Merry Christmas and Happy new year to all the staff and TA's. It's definitely the best course that I have taken. 8

Generating Speech Output

Anyone else notice the format of the asymptotic variance? 2

[Polling] What are your answers for this page?
I think I botched this part of the exam! hahaha oh well. What is your approach in answering this problem? 1

[Staff]Probability
For those of us who have taken other micromasters classes like probability, are we permitted to go back review pertinant course materials to aid in answering questions? 4

Not able to submit answer to "Asymptotic Covariance Matrix" question
Staff, I am not able to submit my answer to the above question. I am getting an error that "Tensor expressions have been forbidden in this entry." I have saved my answer... 2

[STAFF] Please clarify
In the second multi-choice menu (on this page), does the second option “No” mean “Not always” or does it mean “Never”? In the second case, does the option “Not enoug... 2

part 5)
If we want to enter a column vector in part 5 are we supposed to use trans in the answer? or do we just used the brackets and assume the grader understands? The instru... 3

Assumptions in Independence of X and Y, Part 2
[Edited by staff to remove exam content] 2

Problem with rendering
Hi, I'm seeing the following message in the middle of part 5. **Could not format HTML for problem. Contact course staff in the discussion forum for assistance.** Can so... 3

Generating Speech Output