Dauphine | PSL
UNIVERSITÉ PARIS

MIDO - L3 Math. Appliquées 2022–2023

Statistical modelling

Examen final du 11 Janvier 2023

DURÉE 2H00 – DOCUMENTS ET CALCULATRICE NON AUTORISÉS

French – English Lexicon

- *i.i.d.* : *independent and identically distributed*
- échantillon : *sample*
- fonction de répartition : *cumulative distribution function*
- fonction de densité : *probability distribution function*
- fonction génératrice des moments : *moment-generating function*
- famille exponentielle : *exponential family*
- espace naturel des paramètres : *natural parameter space*
- vraisemblance : *likelihood*
- statistique exhaustive : *sufficient statistic*
- statistique libre : *ancillary statistic*
- statistique complète : *complete statistic*

Exercise 1

...... /**9**

For the following statements, give the correct answer(s). Incorrect answers and missing justification return zero point while incomplete answers gain partial points.

1. Let $(X_n)_{n\in\mathbb{N}^*}$ be a sequence of independent discrete random variables such that

$$\mathbb{P}[X_n = 0] = \frac{n-1}{n} \quad \text{and} \quad \mathbb{P}[X_n = \sqrt{n}] = \frac{1}{n}.$$

Then, when n goes to $+\infty$,

(a) the sequence converges in L^1 (convergence in mean),

(b) the sequence converges in L^2 (convergence in quadratic mean),

(c) for any continuous function g, $\mathbb{E}[g(X_n)]$ converges to 0,

(d) the sequence converges in distribution,

(e) the sequence does not converge at all.

(a, d) For any continuous and bounded function g, we have

$$\mathbb{E}\left[g(X_n)\right] = \frac{n-1}{n} g(0) + \frac{1}{n} g(\sqrt{n}).$$

Since g is bounded, the second term in the above sum converges to 0 when n goes to $+\infty$ and thus

$$\mathbb{E}\left[g(X_n)\right] \xrightarrow[n\to+\infty]{\text{a.s.}} g(0).$$

Then, (X_n) converges in distribution to 0. Moreover, if it converges in L^p, $p \in \mathbb{N}^*$, it is necessarily to 0. We have $\mathbb{E}[X_n] = 1/\sqrt{n}$ and $\mathbb{E}\left[X_n^2\right] = 1$. Thus, X_n converges in L^1 to 0, but not in L^2.

2. Consider the exponential family associated to the Bernoulli distribution with unknwon parameter $p \in (0,1)$. The moment generating function of natural statistic $T(x) = x$ of the family is defined for $t \in \Theta \subseteq \mathbb{R}$ by

(a) $(1-p)\big/(1-p-t)$ (b) $1 + p(\exp(t) - 1)$ (c) $(1-p-t)\big/(1-p)$ (d) $1\big/[1 + p(\exp(t) - 1)]$

(b) To get the canonical form we set $\theta = \log(p) - \log(1-p)$. The canonical form is then

$$f(x \mid \theta) = \frac{1}{1+\exp(\theta)} \exp(\theta x), \quad \theta \in \mathbb{R}.$$

Then the moment generating function is defined for any $t \in \mathbb{R}$ by

$$M(t) = \frac{1+\exp(\theta+t)}{1+\exp(\theta)} = \frac{1+\exp[\log(p)-\log(1-p)+t]}{1+\exp[\log(p)-\log(1-p)]} = 1 + p(\exp(t)-1).$$

3. Consider the density (with respect to the Lebesgue measure on $\mathbb{R}_+^*$) parametrised by an **unknown** $(k,\lambda) \in \mathbb{N}^* \times \mathbb{R}_+^*$ and defined by

$$f(x \mid k,\lambda) = \frac{\lambda^k x^{k-1} \exp(-\lambda x)}{(k-1)!} \mathbb{1}_{x>0}.$$

(a) It constitutes a minimal and canonical exponential family.

(b) It constitutes a minimal exponential family but is not in a canonical form.

(c) It constitutes an exponential family that is neither minimal nor canonical.

(d) None of the other answers.

(a) The density writes as

$$f(x \mid k,\lambda) = \frac{\lambda^k}{(k-1)!} \frac{1}{x} \mathbb{1}_{x\geq 0} \exp\left[k \log(x) - \lambda x\right].$$

Then it constitutes a canonical exponential family with natural parameter (k,λ) and natural statistic $T(x) = (\log(x), -x)$. Moreover for $(\alpha_1,\alpha_2) \in \mathbb{R}^* \times \mathbb{R}^*$ and $c \in \mathbb{R}$, the set $\{x \in \mathbb{R}_+ ; \alpha_1 \log(x) - \alpha_2 x - c = 0\}$ contains at most 2 elements (maximal number of intersections between an affine function and $x \mapsto \log(x)$) and hence has measure zero (null set) for the Lebesgue measure. The family is minimal.

4. We run an experiment where we measure how much time n different customers spend on a specific page of a website. Our observations $x_1, \ldots, x_n$ are stored in a vector `x`. We assume that the underlying statistical model is a Gamma distribution with parameter (α,β). Which one among the following R command lines does return the first quartile of the sample?

(a) `rgamma(0.25, 1, 2)`

(b) `pgamma(0.25, 1, 2)`

(c) `dgamma(0.25, 1, 2)`

(d) `qgamma(0.25, 1, 2)`

(e) `quantile(0.25, 1, 2)`

(f) `quantile(x, 0.25)`

(f) In order to get the empirical quantiles of a sample we use the function `quantile`. The first argument is the sample, followed by the order of the quantiles we are interested in.

5. Let X be a random variable with density, parametrised by $\lambda \in \mathbb{R}_+^*$, with respect to the composition of the counting measure on $\mathbb{N}$ and the Lebesgue measure on $\mathbb{R}_+^*$:

$$f_X(x) = \begin{cases} \frac{\lambda^x \exp(-\lambda)}{2(x!)}, & \text{if} \quad x \in \mathbb{N}, \\ \frac{\lambda}{2}\exp(-\lambda x), & \text{otherwise.} \end{cases}$$

The likelihood for the sample $(1,1,2,2,2,x_1,\dots,x_n)$, with $x_1,\dots,x_n \notin \mathbb{N}$ is

(a) $\dfrac{\lambda^n}{2^{n+5}} \exp\left(-\lambda \sum_{i=1}^n x_i\right)$

(b) $\dfrac{\lambda^{n+8}}{2^{n+8}} \exp\left[-\lambda\left(\sum_{i=1}^n x_i + 5\right)\right]$

(c) $\dfrac{\lambda^8 \exp(-5\lambda)}{2^{n+8}}$

(d) $\dfrac{\lambda^{n+3}}{2^{n+6}} \exp\left[-\lambda\left(\sum_{i=1}^n x_i + 2\right)\right]$

(b) The likelihood is given by

$$\left[\frac{\lambda}{2}\exp(-\lambda)\right]^2 \left[\frac{\lambda^2 \exp(-\lambda)}{4}\right]^3 \prod_{i=1}^n \frac{\lambda}{2}\exp(-\lambda x_i) = \frac{\lambda^{n+8}}{2^{n+8}} \exp\left[-\lambda\left(\sum_{i=1}^n x_i + 5\right)\right].$$

6. Consider X distributed according to the Binomial distribution $\mathscr{B}(n,p)$, $n \in \mathbb{N}^*$ **known** and $p \in (0,1)$ **unknown**. If we denote θ the parameter of the canonical form of this exponential family, $I(p)$ and $I(\theta)$ the Fisher information contained in X for p and θ respectively, we have

(a) $I(p) = n/[p(1-p)]$

(b) $I(p) = 1/[p(1-p)]$

(c) $I(\theta) = ne^{-\theta}\left(1+e^{\theta}\right)^2$

(d) $I(\theta) = ne^{\theta}/\left(1+e^{\theta}\right)^2$

(a, d) The density $f(\cdot \mid n,p)$ of the Binomial distribution is twice differentiable with respect to p on $(0,1)$ and

$$\frac{\mathrm{d}}{\mathrm{d}p} f(x \mid n,p) = \frac{x}{p} - \frac{n-x}{1-p} \quad \text{and} \quad \frac{\mathrm{d}^2}{\mathrm{d}p^2} f(x \mid n,p) = -\frac{x}{p^2} - \frac{n-x}{(1-p)^2}.$$

Using that $\mathbb{E}_p[X] = np$, we then have

$$I(p) = -\mathbb{E}_p\left[\frac{\mathrm{d}^2}{\mathrm{d}p^2} f(X \mid n,p)\right] = \frac{np}{p^2} - \frac{n-np}{(1-p)^2} = \frac{n}{p(1-p)}.$$

The parameter of the canonical form is $\theta = \log(p) - \log(1-p)$, that is $p = \exp(\theta)/[1+\exp(\theta)] := \psi(\theta)$. We then have

$$I(\theta) = \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\psi(\theta)\right)^2 I(p) = \frac{\exp(2\theta)}{[1+\exp(\theta)]^4} \frac{n[1+\exp(\theta)]^2}{\exp(\theta)} = \frac{n\exp(\theta)}{[1+\exp(\theta)]^2}.$$

7. Consider a regular and minimal exponential family with natural statistic $T(\cdot)$ and density $f(\cdot \mid \theta)$, $\theta \in \Theta \subseteq \mathbb{R}$. For $X_1,\dots,X_n$ *i.i.d.* random variables distributed according to $f(\cdot \mid \theta)$, we set $S = \sum_{i=1}^n T(X_i)$.

(a) Any bijective transform of S is sufficient for θ.

(b) S is minimal sufficient for θ.

(c) Any sufficient statistic for θ that is a function of S is minimal sufficient for θ.

(d) None of the other answers.

(a, b, c) S is a sufficient statistic for θ. Thus any bijective transform of S is sufficient for θ. For a minimal representation, S is minimal sufficient. Thus S is a function of any other sufficient statistic. Therefore any sufficient statistic R that is a function of S is also a function of any other sufficient statistic. Hence R is also minimal sufficient.

8. Let $X_1, \dots, X_n$ be *i.i.d.* random variables distributed according to the normal distribution $\mathcal{N}(\mu, 1)$ and denote

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad X_{(1)} = \min(X_1, \dots, X_n) \quad \text{and} \quad X_{(n)} = \max(X_1, \dots, X_n).$$

(a) $X_{(n)}$ - $X_{(1)}$ is independent of $\overline{X}_n$.

(b) $\left(X_{(1)}, X_{(n)}\right)$ is not a complete statistic.

(c) $\left(X_{(1)}, X_{(n)}\right)$ is a sufficient statistic for μ.

(d) $\left(X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n\right)$ is independent of $\overline{X}_n$.

(e) None of the other answers.

(a, b, d) $X_{(n)}$ - $X_{(1)}$ and $(X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n)$ are ancillary statistic. Moreover $\overline{X}_n$ is a complete and sufficient statistic for μ (result on natural statistic associated to an exeponential family). It follows from Basu's theorem that both $X_{(n)}$ - $X_{(1)}$ and $(X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n)$ are independent from $\overline{X}_n$.

$X_{(n)}$ - $X_{(1)}$ is an ancillary statistic that is not constant almost surely and such that $\mathbb{E}_\mu\left[X_{(n)} - X_{(1)}\right] = c < \infty$ is independent of μ. Thus, for the function $\phi : (x, y) \mapsto y - x - c$, we have

$$\mathbb{E}_\mu\left[\phi\left(X_{(1)}, X_{(n)}\right)\right] = 0, \quad \forall \mu \in \mathbb{R}.$$

But $\mathbb{P}_\mu\left[\phi\left(X_{(1)}, X_{(n)}\right) = 0\right] = \mathbb{P}_\mu\left[X_{(n)} - X_{(1)} = c\right] \neq 1$ since $X_{(n)}$ - $X_{(1)}$ is not constant almost surely. Therefore $\left(X_{(1)}, X_{(n)}\right)$ is not a complete statistic.

Exercise 2

...... /**12**

Given $a \in \mathbb{R}$, the Lévy distribution of scale parameter $b \in \Theta_0 \subseteq \mathbb{R}_+^*$ admits a density with respect to the Lebesgue measure on $(a, +\infty)$

$$f(x \mid b) = \frac{1}{x-a}\sqrt{\frac{b}{2\pi(x-a)}}\exp\left(-\frac{b}{2x-2a}\right)\mathbb{1}_{x>a}.$$

In this exercise we consider $X_1, \dots, X_n$, $n > 2$, *i.i.d.* random variables distributed according to $f(\cdot \mid b)$ for a **known** parameter a and an **unknown** scale parameter b.

1. Show that $f(\cdot \mid b)$ can define an exponential family with a natural statistic $T(\cdot)$. Precise its canonical form and its natural parameter space. Is the family regular and minimal?

The density writes as

$$f(x \mid b) = c(b)h(x)\exp(\eta(b)T(x)), \quad \text{where} \quad \begin{cases} c(b) = \sqrt{\frac{b}{2\pi}}, & h(x) = \frac{1}{(x-a)\sqrt{x-a}}\mathbb{1}_{x \geq a}, \\ \eta(b) = b, & T(x) = -\frac{1}{2x-2a}. \end{cases}$$

The family is already in its canonical form and its natural parameter space is

$$\Theta = \{b \in \mathbb{R} \mid c(b) > 0\} = \mathbb{R}*_+.$$

The family is regular since Θ is an open set, and minimal, since it is a one dimensional family.

2. Show that there is a **unique** maximum likelihood estimator of b defined as

$$\widehat{b}_n = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{X_i - a}\right)^{-1}$$

The log-likelihood writes as

$$\ell(b \mid x_1,\dots,x_n) = \sum_{i=1}^{n}\log f(x_i \mid b) = \frac{n}{2}\log(b) - b\sum_{i=1}^{n}\frac{1}{2(x_i-a)} - n\log(2\pi) - \frac{3}{2}\sum_{i=1}^{n}\log(x_i - a).$$

It is differentiable on $\Theta = \mathbb{R}_+^*$ and

$$\frac{\partial}{\partial b}\ell(b \mid x_1,\dots,x_n) = \frac{n}{2b} - \sum_{i=1}^{n}\frac{1}{2(x_i-a)}.$$

The log-likelihood has a unique critical point in Θ given by

$$\widehat{b}_n = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i - a}\right)^{-1}$$

Since $f(\cdot \mid b)$ defines a regular exponential family, the critical point is the unique maximum likelihood estimator of b.

3. Show that $\widehat{b}_n \xrightarrow[n\to+\infty]{\mathbb{P}} b$.

We have

$$\widehat{b}_n = \left(-\frac{2}{n}\sum_{i=1}^{n}T(X_i)\right)^{-1}.$$

$T(X_1),\dots,T(X_n)$ is a sequence of *i.i.d.* (as measurable transform of *i.i.d.* random variables) and integrable (comparative growth in a and Riemann criterion in $+\infty$) random variables. The Law of Large Numbers gives

$$\frac{1}{n}\sum_{i=1}^{n}T(X_i) \xrightarrow[n\to+\infty]{\mathbb{P}} \mathbb{E}[T(X_1)].$$

The first moment of the natural statistic $T(X_1)$ is given by

$$\mathbb{E}[T(X_1)] = -\frac{\mathrm{d}}{\mathrm{d}b}\log c(b) = -\frac{1}{2b}. \tag{1}$$

The continuity of $g : x \mapsto (-2x)^{-1}$ on $\mathbb{R}_+^*$ gives that

$$\widehat{b}_n = g\left(\frac{1}{n}\sum_{i=1}^{n}T(X_i)\right) \xrightarrow[n\to+\infty]{\mathbb{P}} g\left(\mathbb{E}[T(X_1)]\right) = b.$$

4. Show that

$$\sqrt{\frac{n}{2}}\left(\frac{\widehat{b}_n - b}{\widehat{b}_n}\right) \xrightarrow[n\to\infty]{d} \mathcal{N}(0,1).$$

Since the exponential family defined by $f(\cdot \mid b)$ is regular, the fundamental theorem of statistic gives

$$\sqrt{n}(\widehat{b}_n - b) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, I_{X_1}(b)^{-1}\right).$$

The log-likelihood is twice differentiable and

$$\frac{\partial^2}{\partial b^2}\ell(b \mid x_1) = -\frac{1}{2b^2}$$

The Fisher information contained in X_1 on b is then

$$I_{X_1}(b) = -\mathbb{E}_b\left[-\frac{1}{2b^2}\right] = \frac{1}{2b^2}.$$

We thus have

$$\sqrt{n}(\widehat{b}_n - b) \xrightarrow[n\to\infty]{d} \mathcal{N}\left(0, 2b^2\right).$$

Using the previous question and the continuity of $x \mapsto x^{-1}$ at $b > 0$, we get

$$\frac{1}{\sqrt{2}\widehat{b}_n} \xrightarrow[n\to+\infty]{\mathbb{P}} \frac{1}{\sqrt{2}b}.$$

It follows from Slutsky's theorem that

$$\frac{\sqrt{n}}{\sqrt{2}\widehat{b}_n}\left(\widehat{b}_n - b\right) \xrightarrow[n\to+\infty]{d} \frac{1}{\sqrt{2}b}\mathcal{N}(0, 2b^2) \equiv \mathcal{N}(0,1).$$

5. Prove that the bias of the estimator $\widehat{b}_n$ is strictly positive.

Since the function $x \mapsto 1/x$ is strictly convex, the Jensen inequality and equation (1) from question 2. lead to

$$\mathbb{E}\left[\widehat{b}_n\right] > \frac{1}{\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \frac{1}{x_i - a}\right]} = \frac{1}{-2\mathbb{E}[T(X_1)]} = b.$$

The bias is then strictly positive.

6. Given a vector of observations x, write a R code that computes a non-parametric bootstrap estimation of the bias for k bootstrap samples.

```r
b_ref <- 1 / mean(1/(x - a))
# --- Solution with a for loop
b_star <- rep(0, k)
for (i in seq_len(k)) {
  b_star[i] <-  1 / mean(1/(sample(x, length(x), TRUE) - a))
}
bias <- mean(b_star) - b_ref
# --- Solution with apply function
mean(apply(matrix(sample(x, k * length(x), TRUE), length(x)), 2,
  function(x, a) {1 / mean(1/(x - a))}, a = a)) - b_ref
```

7. Denote $b_1^\star, \dots, b_k^\star$ the estimations of b we got with the previous bootstrap procedure. Give the definition of the empirical boostrap confidence interval on b with $1-\alpha$ confidence level, $\alpha \in (0,1)$. Choose the R code that computes this interval for a 90% confidence level, where `b_ref` $= \widehat{b}_n$ and `b_star` $= (b_1^\star, \dots, b_k^\star)$.

(a) `quantile(b_star, c(.05, .95))`

(b) `b_ref - quantile(b_star - b_ref, c(.95, .05))`

(c) `quantile(b_star, c(.025, .975))`

(d) `b_ref - quantile(b_star - b_ref, c(.975, .025))`

The empirical boostrap confidence interval on b with $1-\alpha$ confidence level is the interval $[\widehat{b}_n - q_{1-\alpha/2}, \widehat{b}_n - q_{\alpha/2}]$ where q_β denotes the quantile of order $\beta \in (0,1)$ of $(\beta_1^\star - \widehat{b}_n, \dots, \beta_k^\star - \widehat{b}_n)$. For $\alpha = 10\%$ the corresponding code is then (b).

8. Show that if X_1 is distributed according to $f(\cdot \mid b)$ then $Y_1 = (2X_1 - 2a)^{-1}$ admits a density $g(\cdot \mid b)$ with respect to the Lebesgue measure on $\mathbb{R}_+^*$ given by

$$g(y \mid b) = \frac{\sqrt{b}}{\sqrt{\pi y}} \exp(-by) \mathbb{1}_{y>0}.$$

$\phi : x \mapsto (2x-2a)^{-1}$ is a $\mathscr{C}^1$-diffeomorphism between $(a, +\infty)$ and $\mathbb{R}_+^*$. The density of Y_1 is then

$$g(y \mid b) = f\left(\phi^{-1}(y) \mid b\right) \left| \frac{\mathrm{d}}{\mathrm{d}y} \phi^{-1}(y) \right| = f\left(\frac{1}{2y} + a \,\middle|\, b\right) \times \frac{1}{2y^2} = \sqrt{\frac{b}{\pi}} \frac{1}{\sqrt{y}} \exp\left(-by\right) \mathbb{1}_{y>0}.$$

9. Admit the following result : if $Y_1, \dots, Y_n$ are *i.i.d.* random variables distributed according $g(\cdot \mid b)$ then $n/(Y_1 + \dots + Y_n)$ has a density $d(\cdot \mid b)$ with respect to the Lebesgue measure on $\mathbb{R}_+^*$ given by

$$d(x \mid b) = \frac{(nb)^{n/2}}{\Gamma(n/2)} x^{-\frac{n}{2}-1} \exp\left(-\frac{nb}{x}\right) \mathbb{1}_{x>0}.$$

Show that the following estimator is an unbiased estimator of b :

$$\widehat{\beta}_n = \frac{n-2}{n} \widehat{b}_n.$$

Hint. You can use without justification that for all $\alpha > 1$, $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ and

$$\int_0^{+\infty} \frac{(nb)^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(-\frac{nb}{x}\right) \mathrm{d}x = 1.$$

We have

$$\mathbb{E}_b[\widehat{\beta}_n] = \frac{n-2}{n} \mathbb{E}_b[\widehat{b}_n] = \frac{n-2}{n} \mathbb{E}_b\left[\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{X_i - a}\right)^{-1}\right] = \frac{n-2}{2n} \mathbb{E}_b\left[\frac{n}{\sum_{i=1}^n Y_i}\right].$$

Using the density of $n/(Y_1 + \dots + Y_n)$, we have

$$\mathbb{E}_b\left[\frac{n}{\sum_{i=1}^n Y_i}\right] = \int_0^\infty x d(x \mid b) \mathrm{d}x = \int_0^\infty \frac{\sqrt{nb}^n}{\Gamma(n/2)} x^{-\frac{n}{2}} \exp\left(-\frac{nb}{x}\right) \mathrm{d}x = \frac{nb\Gamma(n/2-1)}{\Gamma(n/2)} \int_0^\infty \frac{(nb)^{\frac{n}{2}-1}}{\Gamma(n/2-1)} x^{-\frac{n}{2}} \exp\left(-\frac{nb}{x}\right) \mathrm{d}x$$

Using the hint ($n/2 > 1$ for $n > 2$), we get that the integral is equal to one and $\Gamma(n/2) = (n/2-1)\Gamma(n/2-1)$. Thus,

$$\mathbb{E}_b\left[\frac{n}{\sum_{i=1}^n Y_i}\right] = \frac{2nb}{n-2},$$

and the result follows.

10. Show that

$$\mathbb{V}\mathrm{ar}_b\left[\widehat{b}_n\right] \geq 2\left(\frac{nb}{n-2}\right)^2.$$

We deduce from the previous question that the bias of $\widehat{b}_n$ is

$$\mathbb{E}_b\left[\widehat{b}_n\right] - b = \frac{2}{n-2}b.$$

The Cramér-Rao bound is then

$$\mathbb{V}\mathrm{ar}_b\left[\widehat{b}_n\right] \geq \left(1 + \frac{2}{n-2}\right)^2 \frac{1}{I_{X_1}(b)} = \left(\frac{n}{n-2}\right)^2 2b^2.$$

11. Show that $\widehat{\beta}_n$ is the unique uniformly minimum variance unbiased estimator.

$\widehat{\beta}_n$ is an unbiased estimator of b. Moreover, we have

$$\widehat{b}_n = \left(-\frac{2}{n}\sum_{i=1}^{n} T(X_i)\right)^{-1}.$$

However $T(X)$ is the natural statistic associated to a regular exponential family.Then $S(X_1,\ldots,X_n) = \sum_{i=1}^{n} T(X_i)$ is a complete and sufficient statistic for b. Since $\widehat{b}_n$ is a bijective transform of $S(X_1,\ldots,X_n)$, it is also a complete and sufficient statistic for b. The Lehmann-Scheffé theorem then states that $\mathbb{E}[\widehat{\beta}_n \mid \widehat{b}_n]$ is the unique uniformly minimum variance unbiased estimator. But $\widehat{\beta}_n$ is a measurable function of $\widehat{b}_n$. Then $\mathbb{E}[\widehat{\beta}_n \mid \widehat{b}_n] = \widehat{\beta}_n$.