10. Worked Example Part I

We explore the use of proportionality notation in the process of computing the posterior distribution for a parameter of interest. For this problem, we are given the following information:

- a prior distribution for the parameter λ
- n independent and identically distributed observations $X_1, X_2, \ldots, X_n \in \mathbb{R}$
- the conditional likelihood function $L\left(X_i|\lambda\right)$, assumed to be the same across all observations

Our goal is to use the Bayesian approach to describe the posterior distribution, up to a constant of proportionality. The parameter of interest is λ.

Components of the Bayesian Approach

1/1 point (graded)

Consider Bayes' formula as discussed from the lecture. Which of the following pieces of information are definitely necessary in order to use Bayes' formula to compute the posterior? (Choose all that apply.)

- [ ] The mean of the n observations
- [ ] The Fisher information of the prior distribution $\pi\left(\lambda\right)$
- [x] The likelihood function $L\left(X_1, X_2, \ldots, X_n|\lambda\right)$ of the observations ✔
- [x] The value of the prior distribution $\pi\left(\lambda\right)$ at every point where it is defined. ✔

✔

Solution:

According to Bayes' rule, the posterior distribution (up to a constant of proportionality) is computed by multiplying the prior and posterior distributions taken as a function of the parameter λ. As a result, we need the full distribution for $\pi\left(\lambda\right)$ as well as the liklihood function $L\left(X_1, X_2, \ldots, X_n|\lambda\right)$.

The other two choices, "The mean of the n observations" and "Fisher information of the prior distribution $\pi\left(\lambda\right)$ are not used in the computation of the posterior distribution in the general case. They may, however, show up as part of the likelihood function, after simplification, in particular cases.

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Prior Distribution in Proportionality Notation

4/4 points (graded)

The prior λ is distributed according to **Exp** (a) $(a > 0)$. Write the probability distribution function $\pi\left(\lambda\right)$, in terms of λ and a. Do not simplify.

$\pi\left(\lambda\right) =$

a*exp(-a*lambda) ✔ **Answer:** a*e^(-a*lambda)

$a \cdot \exp(-a \cdot \lambda)$

STANDARD NOTATION

Our expression for $\pi(\lambda)$ uses two variables.

Which one may be ignored (in the context of the problem) if it is used as an outermost multiplier in the expression written in proportional notation?

- (selected) a ✔
- λ
- Both a and λ

Write the simplified expression for $\pi(\lambda)$, ignoring the parameter(s) chosen earlier when used as an outside multiplier.

Hint: Your resulting expression for $\pi(\lambda)$ would satisfy $\pi(0) = 1$ regardless of the value of a.

$\pi(\lambda) =$

exp(-a*lambda) ✔ **Answer:** e^(-a*lambda)

$\exp(-a \cdot \lambda)$

STANDARD NOTATION

Suppose that instead of it being an outermost multiplier, it plays a different role in the formula. Which of the following statements are true?

- [ ] It may still be ignored if the parameter is added to the expression.
- [x] It may still be ignored if the (nonzero) parameter is divided from the expression. ✔
- [ ] It may still be ignored if the expression is taken to the power of the parameter.

✔

Solution:

The definition of an exponential distribution gives an unnormalized $\pi(\lambda) = ae^{-a\lambda}$.

Taken as a function of λ up to a constant of proportionality, we could ignore the a in front and are thus left with $e^{-a\lambda}$. The variable λ may not be dropped because this is our parameter of interest. Thus, if our expression for the PMF depends on it in the original formulation, it must still depend on the parameter in any form simplified by the proportionality notation. Simply dropping the outside a gives $e^{-a\lambda}$. We note that this satisfies the scaling condition $\pi(0) = 1$, so this is the the desired simplification.

Dividing by a quantity that does not depend on our parameter of interest λ is the same as multiplying by the said quantity's reciprocal, so an expression not dependent on the parameter can still be ignored if the expression appears in the denominator. The other two operations, adding and raising to a power, does not preserve the ratio between $\pi(\lambda)$ for different λ values, so we are unable to ignore these in a simplification through the proportionality notation.

Submit You have used 2 of 3 attempts

Answers are displayed within the problem

Likelihood

2/2 points (graded)

We are given the additional information that conditional on the parameter of interest λ, our observations $X_1, X_2, \ldots, X_n$ are independently and identically distributed according to the probability distribution $\textsf{Poiss}\,(\lambda)$. From here, we compute our likelihood function. Our approach would be to compute a single function of λ $L\,(X_i|\lambda)$, then plug in our data $X_1, X_2, \ldots, X_n$ to compute the overall likelihood $L_n\,(X_1, X_2, \ldots, X_n|\lambda) = L\,(X_1|\lambda)\,L\,(X_2|\lambda)\ldots L\,(X_n|\lambda)$.

In our framework so far, we treat the observations $X_1, \ldots, X_n$ as fixed values, by which we perform Bayesian inference. Hence, $L\,(X_i|\lambda)$ is to be viewed as a function of λ with X_i as a parameter. Thus when using proportionality notation, we only need to consider variations concerning our parameter of interest λ. Compute the likelihood function $L\,(X_1|\lambda)$, using proportionality notation to simplify it such that in your expression, $L\,(X_1|\lambda = 1) = e^{-1}$ regardless of the value of X_1. (Note that this is not necessarily the actual likelihood L.)

Use **X1** for X_1.

$L\,(X_1|\lambda) \propto$

lambda^(X1)*exp(-lambc ✔ **Answer:** e^(-lambda)*lambda^(X1)

$\lambda^{X1} \cdot \exp\,(-\lambda)$

Multiply the expressions $L\,(X_1|\lambda)$, $L\,(X_2|\lambda)$, $\ldots$, $L\,(X_n|\lambda)$ based on the simplified expression for $L\,(X_1|\lambda)$ to get the desired likelihood function $L_n\,(X_1, X_2, \ldots, X_n|\lambda)$.

Use **SumXi** for $\sum_{i=1}^{n} X_i$.

$L_n\,(X_1, X_2, \ldots, X_n|\lambda) \propto$

lambda^(SumXi)*exp(-n ✔ **Answer:** e^(-n*lambda)*lambda^(SumXi)

$\lambda^{SumXi} \cdot \exp\,(-n \cdot \lambda)$

STANDARD NOTATION

Solution:

The Poisson distribution on the variable X_1 with parameter λ has pmf $\frac{e^{-\lambda}\lambda^{X_1}}{X_1!}$. Stripping away the constant multiplier $\frac{1}{X_1!}$ gives the expression $e^{-\lambda}\lambda^{X_1}$. We see that this satisfies the condition $L\,(X_1|\lambda = 1) = e^{-1}$ set by the problem statement, so $\boxed{e^{-\lambda}\lambda^{X_1}}$ is indeed our answer.

The general form for $L\,(X_i|\lambda)$ is $e^{-\lambda}\lambda^{X_i}$. Hence,

$$L_n\,(X_1, X_2, \ldots, X_n|\lambda) = L\,(X_1|\lambda)\,L\,(X_2|\lambda)\ldots L\,(X_n|\lambda)$$

$$= (e^{-\lambda}\lambda^{X_1})\ldots(e^{-\lambda}\lambda^{X_n}) = e^{-n\lambda}\lambda^{X_1+\ldots+X_n}$$

$$= \boxed{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} X_i}}.$$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

IID Assumptions in Likelihood Calculation

0/1 point (graded)

One key aspect that made the above computation simple is that we were allowed to simply multiply the same likelihood functional form across different observations. This rests upon a central assumption of both frequentist and Bayesian inference, that our observations are independent and identically distributed. Which of the following statements are true about removing one of the two i.i.d. assumptions? (Choose all that apply.)

- [ ] If we remove only the assumption that the observations are independent conditional on λ, the formula $L_n(X_1, X_2, \ldots, X_n|\lambda) = L(X_1|\lambda) L(X_2|\lambda) \ldots L(X_n|\lambda)$ will still hold.
- [ ] If we remove only the assumption that the observations are identically distributed, the formula $L_n(X_1, X_2, \ldots, X_n|\lambda) = L(X_1|\lambda) L(X_2|\lambda) \ldots L(X_n|\lambda)$ will still hold. ✔ 独立的事件才能相乘
- [x] If we remove only the assumption that the observations are independent given λ, we are still allowed to use a single function of λ for all of the likelihood expressions $L(X_1|\lambda), L(X_2|\lambda), \ldots, L(X_n|\lambda)$. ✔
- [x] If we remove only the assumption that the observations are identically distributed, we are still allowed to use a single function of λ for all of the likelihood expressions $L(X_1|\lambda), L(X_2|\lambda), \ldots, L(X_n|\lambda)$.

因为不是同分布的，所以每个likelihood function都是不同的

✖

Solution:

We consider the two assumptions in order.

- **"Independent" means that observations are independent conditional on λ**: the product rule for independent events allows us to split the joint likelihood into a product of individual liklihoods. Beyond splitting, however, independence has nothing to do with whether the individual likelihood functions are the same.
- **"Identically Distributed" means that a single function is used for the likelihood expressions**: if the distributions are identical, then the likelihood functions, which come from the PMF of the distribution parametrized by a variable λ, would be the same. Being identically distributed has no connection as to whether the joint likelihood may be split; this, rather, is a property of (conditional) independence.

Submit You have used 2 of 2 attempts

Answers are displayed within the problem

Combining with Bayes' Forula

1/1 point (graded)

According to Bayes' formula, $\pi(\lambda|X_1, X_2, \cdots, X_n) \propto \pi(\lambda) L_n(X_1, X_2, \cdots, X_n|\lambda)$. This yields the posterior distribution up to a constant of proportionality. Multiply the relevant expressions above (use the simplified version with proportionality notation) to compute $\pi(\lambda|X_1, X_2, \cdots, X_n)$.

Use **SumXi** for $\sum_{i=1}^n X_i$.

$\pi(\lambda|X_1, X_2, \cdots, X_n) \propto$

exp(-a*lambda)*lambda ✔ **Answer:** e^(-(a+n)*lambda)*lambda^(SumXi)

$\exp(-a \cdot \lambda) \cdot \lambda^{SumXi} \cdot \exp(-n \cdot \lambda)$

STANDARD NOTATION

Solution:

Using Bayes formula above, we get

$$\pi(\lambda|X_1, X_2, \cdots, X_n) \propto (e^{-a\lambda})(e^{-n\lambda}\lambda^{\sum_{i=1}^n X_i}) = \boxed{e^{-(a+n)\lambda}\lambda^{\sum_{i=1}^n X_i}},$$

which is our posterior distribution.