

Análisis Bayesiano en Turismo

Villa de Leyva, Colombia: conteos, proporciones y decisiones



Salimos del Externado temprano, cuando el campus todavía huele a café y a “yo sí estudié, profe”, pero el grupo ya traía energía de paseo de olla con excusa académica. Valentina intentó imponer orden desde el primer minuto: “**Bueno equipo, esto es una salida académica. Académica. No es paseo de olla.**” La frase duró exactamente tres segundos, porque Brayan Camilo levantó una bolsa gigantesca como evidencia en contra: “**¿Y entonces por qué traje chancletas, carbón y una olla como de ejército?**” Nicolás, cargando una carpa que parecía un apartaestudio, lo dijo con total tranquilidad: “**Porque esto sí es paseo de olla, Valentina. . . solo que con inferencia bayesiana.**” Danna Vanessa, en modo supervivencia, hizo la pregunta correcta: “**¿Ya definieron si la prioridad es la *prior*. . . o el almuerzo?**” Luis Alejandro, con tono de profesor y hambre honesta, zanjó el debate: “**Ambas. Sin prior no hay posterior. . . y sin almuerzo no hay energía para calcular nada.**”

La verdadera razón del viaje, sin embargo, no estaba en la libreta de Valentina sino en la mochila térmica de Brayan: el plan era llegar a Villa de Leyva y montar un **puesto de tamales** cerca de la Plaza Mayor. El problema es que vender tamales es fácil *solo si uno adivina bien la demanda*: si compran pocos, se quedan sin ventas; si compran demasiados, terminan regalándolos o cargándolos todo el día como evidencia de mala planeación. Por eso este taller existe: para usar inferencia bayesiana y decidir con datos **cuántos tamales llevar y dónde venderlos** (en el centro o fuera del centro), con intervalos de credibilidad y predicción que muestren el riesgo de quedarse cortos o pasarse.

David Ricardo venía pegado al celular como si la señal fuera un derecho humano: “**Yo solo quiero saber si hay internet. Si no hay señal, yo no estimo nada.**” Allison

Michelle, feliz viviendo en el caos, lo calmó con una sonrisa: **“Si no hay señal, hacemos MCMC a mano: Me Como Carne... perdón, Markov Chain Monte Carlo.”** Y justo cuando el grupo parecía completo, apareció Jhon Jairo tarde —como evento exógeno— preguntando con inocencia: **“¿Me perdí algo?”** La respuesta fue unánime y dramática: **“¡LA SALIDA!”** Él, sin despeinarse, reveló su aporte a la teoría y a la vida: **“Tranquilos, yo traje lo más importante: el ají.”** Valentina respiró hondo, lo anotó en su libreta y dictó sentencia: **“Variable crítica.”**

En la buseta empezó el experimento real: el choque entre teoría y realidad con temperatura alta y paciencia baja. Brayan Camilo quiso convertir el trayecto en festival: **“Pongo música pa’ que el bus acelere.”** Valentina lo miró con esa mezcla de paciencia y amenaza que solo produce una hoja de Excel mal formateada: **“No. Si aceleras, la varianza de mi paciencia explota.”** Nicolás, inspirado por el calor y el parlante, convirtió el caos en modelo: **“Esto es Gamma–Poisson: la tasa de gritos por minuto sube con el volumen.”** Luis Alejandro intentó rescatar la formalidad: **“Técnicamente, los gritos por intervalo siguen Poisson si asumimos independencia... cosa que Brayan no respeta.”** Danna Vanessa, con honestidad brutal, cerró el asunto: **“Yo solo sé que cuando Brayan canta, mi probabilidad de bajarme del bus tiende a 1.”** David Ricardo, siempre dispuesto a modelar el sufrimiento, preguntó: **“¿Y si el número de paradas inesperadas también lo modelamos como Poisson?”** Allison, sin piedad, remató: **“Conjugado con Gamma, obvio. Aquí todo es conjugado... menos Brayan con el silencio.”**

La salvación llegó en forma de estrategia bayesiana (y carbohidrato). Jhon Jairo, que a pesar de llegar tarde tenía talento para reducir ruido, propuso: **“Si paramos por almojábanas, baja la varianza del sistema.”** Paramos, comimos, y de manera misteriosa el proceso empezó a converger: menos gritos, más paz, y hasta Valentina aceptó que esa intervención tenía evidencia empírica. Ya en Villa de Leyva, el grupo se partió como si la ciudad hubiera ejecutado un algoritmo no supervisado: hostel con “agua caliente a veces” y camping con “incertidumbre honesta”. Y ahí, con ají como variable crítica, quedó oficialmente inaugurado el taller: **los cálculos de abajo no son por deporte**, son para que el puesto de tamales sobreviva al primer día sin quiebra, sin desperdicio y con decisiones justificadas por evidencia.

Objetivos de aprendizaje

Al finalizar, el equipo será capaz de:

- Parametrizar priors conjugadas a partir de media/varianza (Gamma y Beta).
- Construir likelihoods como función del parámetro y **escalarlas** para compararlas visualmente.
- Derivar y resumir posteriors (media, varianza, intervalo creíble 95 %).
- Construir la **predictiva posterior** para decisiones (porciones y ubicación).
- Comunicar con evidencia visual: **intervalos creíbles/predictivos sombreados** y conclusiones justificadas.

Instrucciones y formato (obligatorio)

Instrucciones

Trabajo en equipos de 3.

En cada ejercicio deben responder con:

- (i) **Fórmula/resultado** (prior, posterior, media, IC, etc.).
- (ii) **Interpretación en 2–4 líneas:** *¿qué le hizo el dato a la prior? ¿qué tanto manda la prior?*
- (iii) **Evidencia gráfica** (ver lista G1–G4).

Entrega obligatoria

Para CADA escenario deben entregar 4 gráficas:

- **G1 Prior:** densidad del parámetro (λ o p) con **IC 95 % sombreado**.
- **G2 Likelihood:** como función del parámetro, **escalada a máximo 1**, con el máximo marcado.
- **G3 Posterior:** densidad con **IC 95 % sombreado** y una línea vertical en la media posterior.
- **G4 Predictiva posterior:** distribución del dato futuro con **intervalo predictivo 95 % sombreado** o marcado claramente.

Regla clave: los intervalos deben ir coloreados/sombreados (no solo dos líneas).

Parte A: Gamma–Poisson (conteos en la Plaza Mayor)

Después del viaje en bus y la primera discusión seria sobre almojábanas, el grupo llega a la Plaza Mayor con un problema muy real: si van a hacer paseo de olla, necesitan decidir cuánta comida comprar *para mañana*. El dilema es que el número de turistas no es fijo: algunos días llegan pocos, otros días la plaza se llena, y si se equivocan pasan dos cosas malas: o se quedan cortos (crisis) o compran de más (desperdicio). Para organizar la decisión, el equipo define Y como el número de turistas que llegan en un día y usa el modelo $Y | \lambda \sim \text{Poisson}(\lambda)$, donde λ representa la tasa promedio diaria de turistas (lo que realmente quieren aprender). Como todavía no han contado a nadie, empiezan con una creencia previa sobre λ : los comerciantes dicen que históricamente llegan en promedio 50 turistas por día con varianza 200, y eso se traduce en una prior Gamma que ustedes deben calibrar. Luego aparece el primer dato del día: hoy contaron $y = 55$, y ese conteo funciona como evidencia para “empujar” la creencia inicial mediante la likelihood. La meta de esta parte no es solo calcular una posterior, sino **verla**: por eso deben producir las cuatro gráficas obligatorias (prior con IC 95 % sombreado, likelihood escalada, posterior con IC 95 % sombreado y media marcada, y predictiva posterior con intervalo predictivo 95 % resaltado) y usarlas para explicar una idea central: *¿quién manda más, la prior o el dato?* Para comprobarlo, repetirán todo con una prior alternativa que representa el rumor de

“festival”: una creencia con media 80 pero **poca fuerza** (alta varianza). Al final, con cada escenario (prior informada vs prior débil), construirán la predictiva de $Y_{\text{mañana}} | y$ y cerrarán con una recomendación concreta de porciones: una “eficiente” basada en el valor esperado y otra “prudente” basada en el extremo superior (o percentil 95 %) del intervalo predictivo.

$$Y | \lambda \sim \text{Poisson}(\lambda), \quad \lambda > 0.$$

A1. Prior informada por comerciantes (histórica)

Comerciantes reportan que llegan en promedio $\mu = 50$ turistas/día con varianza $\sigma^2 = 200$. Asuma prior:

$$\lambda \sim \text{Gamma}(\alpha, \beta) \quad (\text{parametrización rate}).$$

1. Calcule α, β usando: $\mathbb{E}[\lambda] = \alpha/\beta$, $\text{Var}(\lambda) = \alpha/\beta^2$.
2. **(G1)** Grafique $p(\lambda)$. **Sombree** el intervalo creíble 95 % de la prior. Marque la media.
3. Interprete: ¿qué significa “prior informada” en términos de dispersión y α, β ?

A2. Dato observado y likelihood

Hoy contaron $y = 55$ turistas.

1. Escriba la likelihood $p(y | \lambda)$ como función de λ .
2. **(G2)** Grafique la likelihood **escalada a máximo 1**. Indique dónde queda el máximo (en λ).

A3. Posterior (conjugación) y comparación

1. Calcule la posterior $\lambda | y$ por conjugación.
2. Reporte: media posterior y un intervalo creíble 95 % para $\lambda | y$.
3. **(G3)** Grafique **prior vs posterior** (dos curvas) y **sombree** el IC 95 % posterior. Marque la media posterior.
4. Interprete: ¿la posterior se movió mucho respecto a la prior? ¿por qué?

A4. Prior débil (festival, pero poca fuerza)

Valentina sospecha festival: $\mathbb{E}[\lambda] = 80$, pero quiere una prior **con alta varianza**.

1. Proponga una Gamma con media 80 y varianza alta. Justifique cómo su elección refleja “poca fuerza”.
2. Repita A2 y A3 con el mismo dato $y = 55$.
3. **(G1–G3)** Deben estar: prior (IC sombreado), likelihood escalada, posterior (IC sombreado) y comparación.
4. Interprete: compare “quién manda” (prior vs datos) en el caso informada vs débil.

A5. Predicción para mañana (decisión de porciones)

1. Para cada escenario (informada vs débil), construya la **predictiva posterior** de $Y_{\text{mañana}} \mid y$.
2. Obtenga el **intervalo predictivo 95 %** para $Y_{\text{mañana}} \mid y$.
3. **(G4)** Grafique la predictiva: pmf (o barras) y **sombree** el intervalo predictivo 95 %.
4. Decisión: recomiende cuántas porciones comprar:
 - Opción “eficiente”: use $\mathbb{E}[Y_{\text{mañana}} \mid y]$.
 - Opción “prudente”: use el percentil 95 % (o extremo superior del intervalo predictivo).

Parte B: Beta–Binomial (proporción en el centro histórico)

Con la comida casi resuelta, aparece el segundo problema: *¿dónde conviene instalar el punto de encuentro y venta?* Villa de Leyva tiene un centro histórico muy atractivo, pero no todos los turistas se quedan allí: algunos se dispersan hacia restaurantes, museos y rutas. Para tomar la decisión, el grupo define p como la proporción real de turistas que permanece en el centro y plantea el modelo $X \mid p \sim \text{Binomial}(n, p)$, donde X es el número de turistas que están en el centro dentro de un conteo de n turistas. Antes de medir, hay opiniones: la experiencia local sugiere que $p \approx 0,40$ con cierta incertidumbre (varianza 0,02), lo cual se traduce en una prior Beta que ustedes deben parametrizar. Después el grupo hace un conteo rápido: en $n = 100$ turistas, $x = 42$ estaban en el centro. Con eso deben construir la likelihood, actualizar la prior para obtener la posterior y, otra vez, **mostrarlo con evidencia visual**: prior con IC 95 % sombreado, likelihood escalada, posterior con IC 95 % sombreado y comparación clara. Luego viene el choque de posturas: Nicolás propone empezar “en blanco” con una Beta(1,1), y el dueño del hostel insiste en que “siempre” es $p \approx 0,80$ con alta seguridad (prior fuerte). El objetivo es que comparen cómo cambian las posteriors (y sus intervalos) según la fuerza de la prior y respondan lo importante con argumentos y gráficas: *¿cuánto empuja el dato 42/100 a cada prior, y cuál escenario queda más confiable para decidir?* Finalmente, como la decisión es operativa, construirán la predictiva posterior para una muestra futura de $m = 50$ turistas y resaltarán el intervalo predictivo 95 % en la gráfica; con eso cierran recomendando si instalar el punto en el centro o no, justificándolo con incertidumbre (no con una sola cifra).

$$X \mid p \sim \text{Binomial}(n, p), \quad 0 < p < 1.$$

B1. Prior basada en experiencia local

Se cree que $p \approx 0,40$ con varianza 0,02. Prior:

$$p \sim \text{Beta}(\alpha, \beta).$$

1. Calcule α, β usando media y varianza de la Beta.
2. **(G1)** Grafique la prior $p(p)$ y **sombree** el IC 95 % prior. Marque la media.

- Interprete: explique $\alpha + \beta$ como **tamaño muestral equivalente** (pseudo-observaciones).

B2. Dato observado, likelihood y posterior

En $n = 100$ turistas observados, $x = 42$ están en el centro.

- Escriba la likelihood $p(x | p)$ como función de p .
- (G2) Grafique la likelihood **escalada a máximo 1** y marque $\hat{p} = x/n$.
- Calcule la posterior $p | x$ por conjugación.
- Reporte media posterior e intervalo creíble 95 %.
- (G3) Grafique prior vs posterior y **sombree** el IC 95 % posterior.
- Interprete: ¿qué tanto se movió desde 0.40 hacia 0.42?

B3. Prior no informativa

Nicolás propone $p \sim \text{Beta}(1, 1)$.

- Calcule la posterior con $n = 100, x = 42$.
- (G3) Compare visualmente **dos posteriors** en la misma figura: (i) prior local y (ii) prior no informativa. Sombrée ambos IC 95 %.
- Conclusión: ¿cuál queda más pegada a 0.42 y por qué?

B4. Prior experta sesgada (conflicto)

El dueño del hostel insiste en $p \approx 0,80$ con **alta fuerza**. Proponga $\alpha + \beta = 100$ (o mayor).

- Construya una Beta con media 0.80 y fuerza $\alpha + \beta = 100$. Reporte α, β .
- Actualice con $x = 42, n = 100$ y obtenga la posterior.
- (G3) Grafique **tres posteriors** (no informativa vs local vs experta) y **sombree** los IC 95 %.
- Interprete usando pseudo-observaciones: ¿quién gana (datos vs experto) y por qué?

B5. Predicción para una muestra futura

Para una nueva muestra de $m = 50$ turistas, defina X_{fut} como número en el centro.

- Para cada escenario (local, no informativa, experta), construya la **predictiva posterior** de $X_{\text{fut}} | x$.
- Obtenga intervalos predictivos 95 %.

3. **(G4)** Grafique la predictiva (pmf) y **sombree** el intervalo predictivo 95 % (o marque el bloque de barras dentro del intervalo).
4. Decisión: ¿instalan el punto de venta en el centro? Justifique con la incertidumbre.

Parte C: Análisis integrador (decisión final)

Entregable de análisis (obligatorio): 1 página (máximo) con:

- Recomendación de porciones para mañana (escenario A) con argumento “eficiente” y “prudente”.
- Recomendación de ubicación (centro vs fuera) usando p y la predictiva (escenario B).
- **Comparación explícita** de cómo cambian las decisiones si se usa prior informada vs prior débil/no informativa vs prior experta.
- Una frase final tipo comité: *“Decidimos X porque con 95 % de credibilidad/predicción...”*

Checklist rápido (para Valentina)

Antes de entregar, verifique:

- Todas las gráficas tienen título, ejes, leyenda y unidades.
- Likelihood está escalada a máximo 1.
- Intervalos 95 % están **sombreados** (no solo dos líneas).
- Media posterior marcada en G3.
- Predictiva incluye intervalo predictivo 95 % (y decisión asociada).
- Análisis integrador (Parte C) en **máximo 1 página**.
- Todo se entrega en un archivo en teams con los nombres de todos.

Rúbrica

Rúbrica (10 puntos)

Criterio	Puntos
G1 Prior (bien parametrizada, IC 95 % sombreado, media marcada)	1.5
G2 Likelihood (función correcta, escalada a 1, máximo marcado)	1.5
G3 Posterior (correcta, IC 95 % sombreado, comparación clara con prior)	2.5
G4 Predictiva (correcta, intervalo predictivo 95 % sombreado/marcado)	2.5
Interpretación (claridad: “qué le hizo el dato a la prior” y “quién manda”)	1.0
Análisis integrador (decisiones justificadas con incertidumbre)	1.0
Total	10.0

Sugerencias técnicas (para sombrear intervalos)

No es obligatorio usar el mismo software, pero sí sombrear intervalos.

- **Python (matplotlib):** usar `fill_between(x, y, where=(x>=L)&(x<=U), alpha=0.3)` para sombrear el IC.
- **R (ggplot2):** usar `geom_area()` o `geom_ribbon()` para sombrear entre L y U .
- **Excel:** graficar la curva y crear una serie auxiliar (igual a la densidad dentro del intervalo y 0 fuera) y formatearla como área con transparencia.