Desagregación de Estimaciones en Áreas Pequeñas un enfoque bayesiano

CEPAL - Unidad de Estadísticas Sociales

Table of contents I

Introducción al pensamiento bayesiano.

Introducción al pensamiento bayesiano.

Modelos de áreas con el enfoque de **Tom**

Y te levantas un día...

- Y te sientes un poco raro, y débil. Vas al médico y te hacen exámenes. Uno de ellos te marca positivo para una enfermedad muy rara que solo afecta al 0.1% de la población.

No son buenas noticias.

- Vas al consultorio del médico y le preguntas qué tan específico es el examen. Te dice que es muy preciso; identifica correctamente al 99% de la gente que tiene la enfermedad.

Y conoces a Thomas...

Esta es la información que tienes:

- $\mathrm{P(E) = 0.001}$
- $\mathrm{P(+|E) = 0.99}$
- $\mathrm{P(-E) = 0.999}$
- $\mathrm{Pr(+| - E) = 0.01}$

Además, por el teorema de probabilidad total

$$\begin{aligned}\mathrm{P(+)} &= \mathrm{Pr(E)Pr(+|E) + Pr(-E)Pr(+| - E)} \\ &= 0.001 * 0.99 + 0.999 * 0.01 \\ &= 0.01098\end{aligned}$$

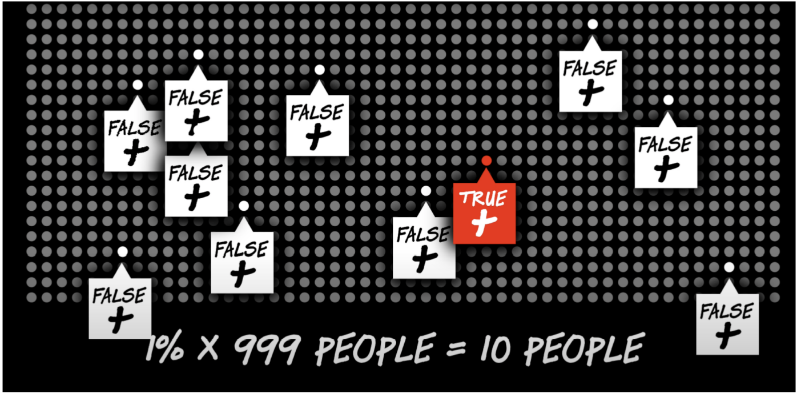
La regla de Bayes afirma lo siguiente:

$$\mathrm{Pr(E|+)} = \frac{\mathrm{Pr(+|E) \times Pr(E)}}{\mathrm{Pr(+)}}$$

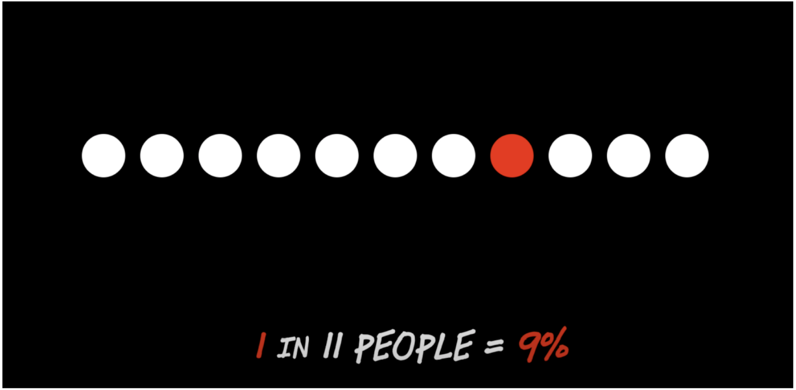
Por lo tanto:

$$\mathrm{Pr(E|+)} = 0.09 \approx 9\%$$

¿cómo funciona?



¿cómo funciona?



Y pides una segunda opinión

- Y esta vez el médico ordena que vuelves a realizarte ese mismo examen... y vuelves a marcar positivo para esa enfermedad.
- **Y vuelves a preguntarte:** *¿cuál es la probabilidad de que tenga esa enfermedad?*

Esta vez, has actualizado tu información sobre $Pr(E)$, pues ya marcaste positivo en un examen

$$Pr(E) = 0.09 \; Y \;\; Pr(-E) = 0.91$$

Por lo tanto:

$$Pr(E \mid ++) = 0.997 \approx 91\%$$

Elementos de la regla de Bayes

En términos de inferencia para θ, es necesario encontrar la distribución de los parámetros condicionada a la observación de los datos. Para este fin, es necesario definir la distribución conjunta de la variable de interés con el vector de parámetros.

$$p(\theta, Y) = p(\theta)p(Y \mid \theta)$$

- La distribución $p(\theta)$ se le conoce con el nombre de distribución previa.
- El término $p(Y \mid \theta)$ es la distribución de muestreo, verosimilitud o distribución de los datos.
- La distribución del vector de parámetros condicionada a los datos observados está dada por

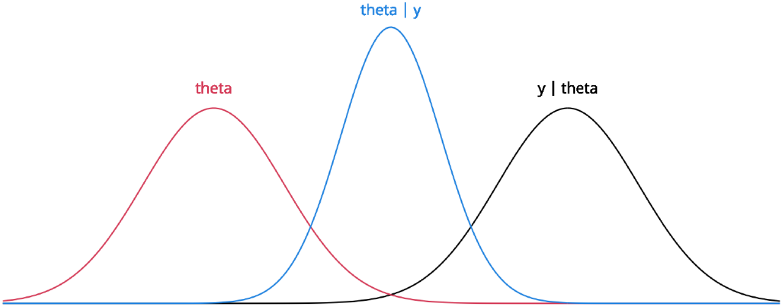
$$p(\theta \mid Y) = \frac{p(\theta, Y)}{p(Y)} = \frac{p(\theta)p(Y \mid \theta)}{p(Y)}$$

Regla de Bayes

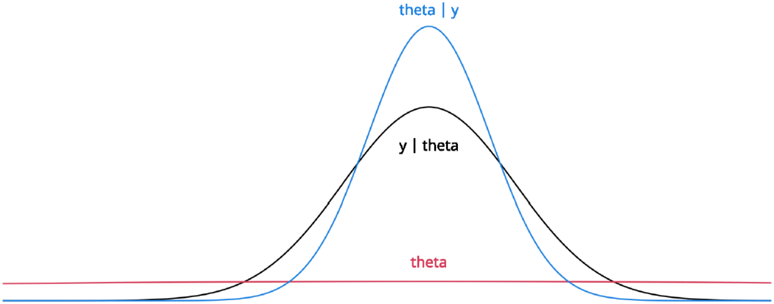
- El término $p(\theta \mid Y)$ se le conoce con el nombre de distribución ***posterior***.
- El denominador no depende del vector de parámetros y considerando a los datos observados como fijos, corresponde a una constante y puede ser obviada. Luego,

$$p(\theta \mid Y) \propto p(Y \mid \theta)p(\theta)$$

Distribución previa informativa para θ



Distribución previa NO informativa para θ



Modelo de área Poisson

Suponga que $Y = \{Y_1, \dots, Y_n\}$ es una muestra aleatoria de variables con distribución Poisson con parámetro θ, la función de distribución conjunta o la función de verosimilitud está dada por

$$p(Y \mid \theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{y_i}}{y_i!} I_{\{0,1,\dots\}}(y_i)$$

$$= \frac{e^{-n\theta}\theta^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!} I_{\{0,1,\dots\}^n}(y_1, \dots, y_n)$$

donde $\{0, 1 \dots\}^n$ denota el producto cartesiano n veces sobre el conjunto $\{0, 1 \dots\}$.

El parámetro θ está restringido al espacio $\Theta = (0, \infty)$.

Distribución previa para θ

- La distribución previa del parámetro θ dada por

$$p(\theta \mid \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \theta^{\alpha-1} e^{-\beta\theta} I_{(0,\infty)}(\theta).$$

- La distribución posterior del parámetro θ está dada por

$$\theta \mid Y \sim Gamma\left(\sum_{i=1}^{n} y_i + \alpha, n + \beta\right)$$

Proceso de estimación en **STAN**

Sea Y el conteo de personas encuestadas que se encuentran por debajo de la línea de pobreza, expresado como una tasa de (X) por cada 100 habitantes, por división administrativa del país.

```r
dataPois <- readRDS("www/00_Intro_bayes/Poisson/dataPoisson.rds")
```

Table 1: Conteno de personas

dam2	n
05002	2
05031	1
05034	1
05045	2
05079	1
05088	6
05093	1
05120	2
05129	1
05142	1

Histrograma con el conteno de personas

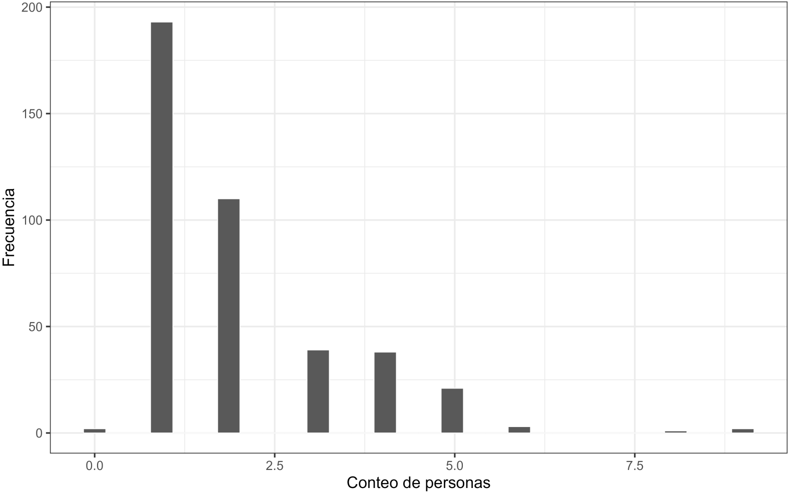


Figure 1: Conteno de personas por división administrativa

Modelo escrito en código STAN

```
data {
  int<lower=0> n;       // Número de áreas geograficas
  int<lower=0> y[n];    // Conteos por area
  real<lower=0> alpha;
  real<lower=0> beta;
}
parameters {
  real<lower=0> theta;
}
model {
  y ~ poisson(theta);
  theta ~ gamma(alpha, beta);
}
generated quantities {
    real ypred[n];                          // vector de longitud n
    for(ii in 1:n){
    ypred[ii] = poisson_rng(theta);
    }
}
```

Preparando datos para código STAN

- Organizando datos para STAN

```
sample_data <- list(n = nrow(dataPois), y = dataPois$n,
                    alpha = 0.001, beta = 0.001)
```

- Ejecutando el código de STAN

```
stan_pois <- "www/00_Intro_bayes/Poisson/03_Poisson.stan"
model_poisson <-
  stan(
    file = stan_pois, data = sample_data,
    warmup = 500,
    iter = 1000,
    verbose = FALSE,   cores = 4
  )
saveRDS(model_poisson,
        "www/00_Intro_bayes/Poisson/model_poisson.rds")
```

Resultados de la estimación del parámetro θ

```
tabla_posi <- summary(model_poisson,
                      pars =c("theta"))$summary
tabla_posi%>%tba()
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
theta	2.03	0.0023	0.0685	1.901	1.982	2.032	2.077	2.166	851.6	1.002

Convergencias de las cadenas el parámetro θ

```
posterior_theta <- as.array(model_poisson, pars = "theta")
p1 <- (mcmc_dens_chains(posterior_theta) +
    mcmc_areas(posterior_theta) ) / mcmc_trace(posterior_theta)
```

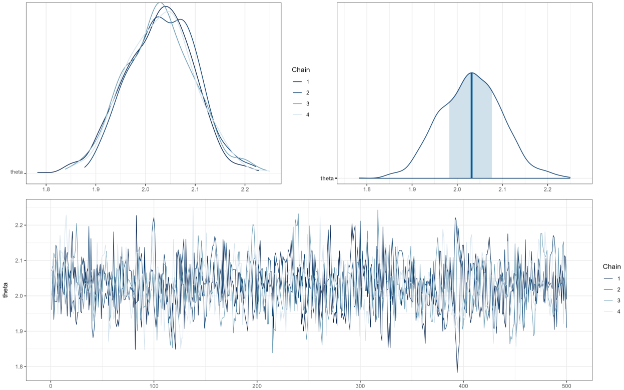


Figure 2: Cadenas para theta

Chequeo predictivo posterior

```
y_pred_B<-as.array(model_poisson,pars ="ypred") %>%
  as_draws_matrix()

rowsrandom<-sample(nrow(y_pred_B),100)

y_pred2<-y_pred_B[rowsrandom,]

p1<- ppc_dens_overlay(y =as.numeric(dataPois$n*100), y_pred2*100)
```

Chequeo predictivo posterior

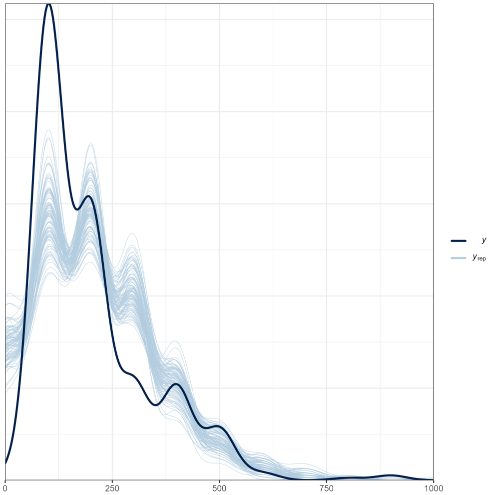


Figure 3: Cadenas para theta

Modelo de unidad: Normal con media y varianza desconocida

- En el modelo normal se considera un conjunto de variables independientes e idénticamente distribuidas $Y_1, \cdots, Y_n \sim N(\theta, \sigma^2)$.
- Cuando se desconocen tanto la media como la varianza de la distribución, se plantean diferentes enfoques para asignar las distribuciones previas para θ y σ^2 según el contexto del problema.

Distribuciones previas para θ y σ^2

Se describen tres posibles suposiciones sobre las distribuciones previas para θ y σ^2, considerando independencia y nivel de informatividad.

- Suponer que la distribución previa $p(\theta)$ es independiente de la distribución previa $p(\sigma^2)$ y que ambas distribuciones son informativas.
- Suponer que la distribución previa $p(\theta)$ es independiente de la distribución previa $p(\sigma^2)$ y que ambas distribuciones son no informativas.
- Suponer que la distribución previa para θ depende de σ^2 y escribirla como $p(\theta \mid \sigma^2)$, mientras que la distribución previa de σ^2 no depende de θ y se puede escribir como $p(\sigma^2)$.

Se establece la distribución previa para el parámetro θ como $\theta \sim Normal(0, 10000)$ y para el parámetro σ^2 como $\sigma^2 \sim IG(0.0001, 0.0001)$.

Definición del modelo normal

- El objetivo del modelo es estimar el ingreso medio de las personas, representado como $\bar{Y}_d = \frac{\sum_{U_d} y_{di}}{N_d}$.
- Se muestra una forma de estimar $\bar{Y}$ mediante el uso de $\hat{y}_{di}$, el cual es el valor esperado de y_{di} bajo una medida de probabilidad inducida por el modelo.
- Finalmente, se presenta la estimación de $\hat{\bar{Y}}_d = \frac{\sum_{U_d} \hat{y}_{di}}{N_d}$.

Proceso de estimación

- Estimar el ingreso medio de las personas, es decir,

$$\bar{Y}_d = \frac{\sum_{U_d} y_{di}}{N_d}$$

donde y_{di} es el ingreso de cada personas. Note que

$$\bar{Y}_d = \frac{\sum_{s_d} y_{di} + \sum_{s_d^c} y_{di}}{N_d}$$

Ahora, el estimador de $\bar{Y}$ esta dado por:

$$\hat{\bar{Y}}_d = \frac{\sum_{s_d} y_{di} + \sum_{s_d^c} \hat{y}_{di}}{N_d}$$

Proceso de estimación

Ahora, es posible suponer que $\hat{y}_{di}$ es la esperanza condicional dado el modelamiento, es decir

$$\hat{y}_{di} = E_{\mathcal{M}}\left(y_{di} \mid x_d, \beta\right)$$

,

donde $\mathcal{M}$ hace referencia a la medida de probabilidad inducida por el modelamiento. Finalmente se tiene que,

$$\hat{\bar{Y}}_d = \frac{\sum_{U_d} \hat{y}_{di}}{N_d}$$

Proceso de estimación en **STAN**

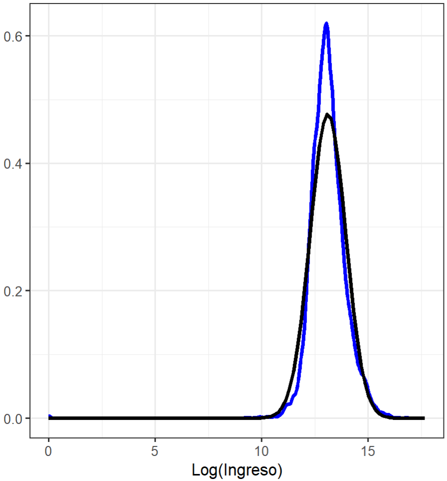
Sea Y el logaritmo del ingreso para una división administrativa del país.

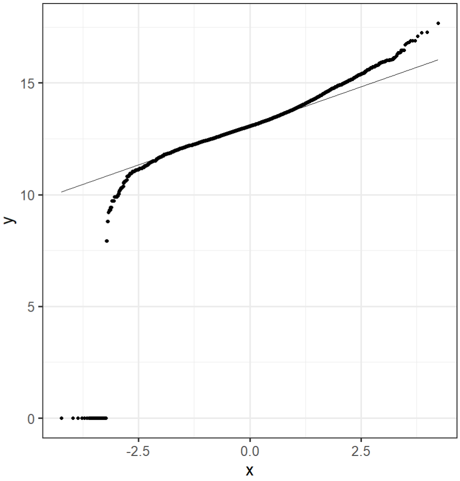
```
dataNormal <- readRDS("www/00_Intro_bayes/Normal/01_dataNormal.rds")
tba(dataNormal %>% head(10), cap = "Logaritmo del ingreso" )
```

Table 2: Logaritmo del ingreso

dam_ee	logIngreso
08	14.37
08	14.37
08	14.37
08	12.02
08	12.02
08	12.02
08	14.05
08	14.05
08	14.05
08	14.05

Analisis gráfico del logaritmo del ingreso.





Modelo escrito en código de STAN

```
data {
  int<lower=0> n;
  real y[n];
}
parameters {
  real sigma;
  real theta;
}
transformed parameters {
  real sigma2;
  sigma2 = pow(sigma, 2);
}
model {
  y ~ normal(theta, sigma);
  theta ~ normal(0, 1000);
  sigma2 ~ inv_gamma(0.001, 0.001);
}
generated quantities {
    real ypred[n];
    for(kk in 1:n){
    ypred[kk] = normal_rng(theta,sigma);
}
}
```

Preparando datos para el código de STAN

- Organizando datos para STAN

```
sample_data <- list(n = nrow(dataNormal),
                    y = dataNormal$logIngreso)
```

- Ejecutando STAN desde R mediante la librería **rstan**

```
NormalMeanVar  <- "www/00_Intro_bayes/Normal/03_NormalMeanVar.stan"
model_NormalMedia <- stan(
  file = NormalMeanVar,
  data = sample_data,
  warmup = 500,
  iter = 1000,
  verbose = FALSE, cores = 4
)
saveRDS(model_NormalMedia,
        "www/00_Intro_bayes/Normal/model_NormalMedia2.rds")
```

Resultados de la estimación del parámetro θ y σ^2 es:

```
tabla_Nor2 <- summary(model_NormalMedia,
        pars = c("theta", "sigma2", "sigma"))$summary

tabla_Nor2 %>% tba()
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_ef
theta	13.1147	1e-04	0.0039	13.1068	13.1120	13.1148	13.1172	13.1224	122
sigma2	0.6986	1e-04	0.0047	0.6894	0.6955	0.6986	0.7016	0.7080	193
sigma	0.8358	1e-04	0.0028	0.8303	0.8340	0.8358	0.8376	0.8415	193

Convergencias de las cadenas el parámetro θ

```
posterior_theta <- as.array(model_NormalMedia, pars = "theta")
(mcmc_dens_chains(posterior_theta) +
    mcmc_areas(posterior_theta) ) /
  mcmc_trace(posterior_theta)
```

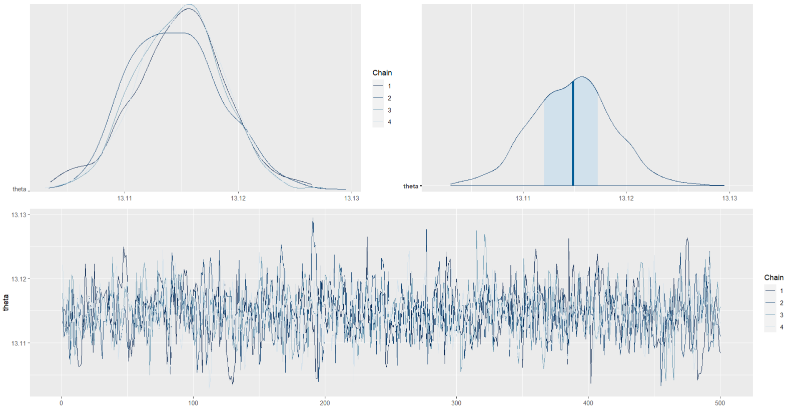


Figure 4: Cadenas para theta

Convergencias de las cadenas el parámetro σ^2

```r
posterior_sigma2 <- as.array(model_NormalMedia, pars = "sigma2")
(mcmc_dens_chains(posterior_sigma2) +
    mcmc_areas(posterior_sigma2) ) /
  mcmc_trace(posterior_sigma2)
```

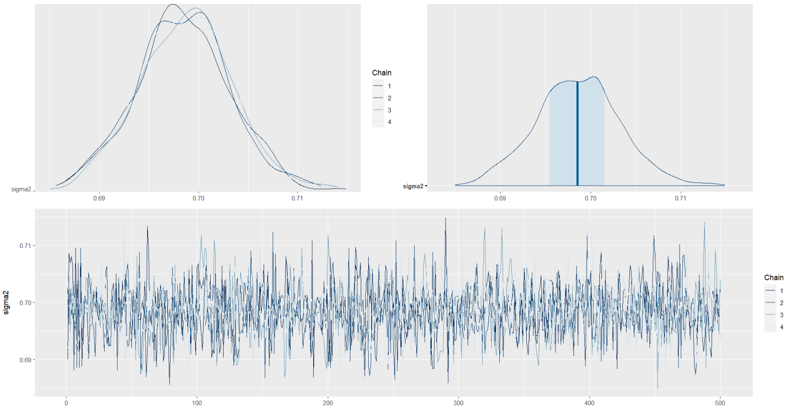


Figure 5: Cadenas para sigma2

Chequeo predictivo posterior del ingreso

```r
y_pred_B <- as.array(model_NormalMedia, pars = "ypred") %>%
  as_draws_matrix()

rowsrandom <- sample(nrow(y_pred_B), 100)

y_pred2 <- y_pred_B[rowsrandom,]

ppc_dens_overlay(
  y = as.numeric(exp(dataNormal$logIngreso) - 1), y_pred2) +
  xlim(0, 5000000)
```

Chequeo predictivo posterior del ingreso

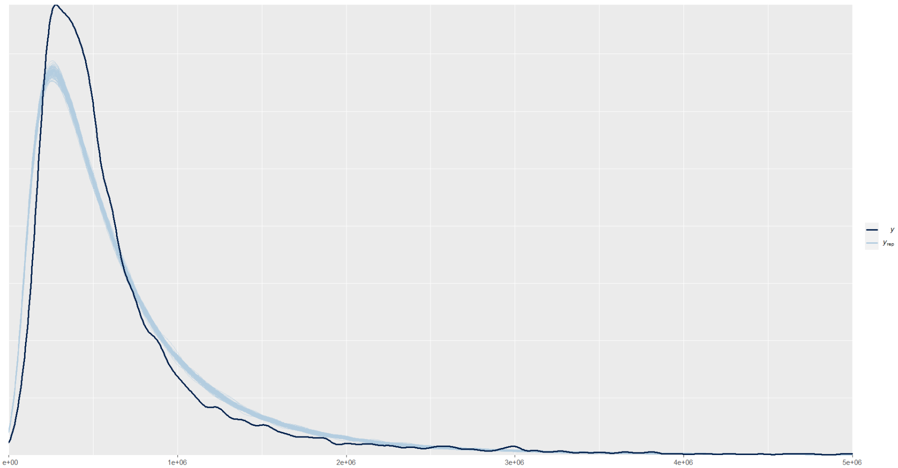


Figure 6: PPC para el ingreso

Modelos lineales.

La regresión lineal es la técnica básica del análisis econométrico. Mediante dicha técnica tratamos de determinar relaciones de dependencia de tipo lineal entre una variable dependiente o endógena, respecto de una o varias variables explicativas o exógenas.

Modelos lineales bayesiano.

En primer lugar, nótese que el interés particular recae en la distribución del vector de n variables aleatorias $Y = (Y_1 \dots, Y_n)'$ condicional a la matriz de variables auxiliares X e indexada por el vector de parámetros de interés $\beta = (\beta_1, \dots, \beta_q)'$ dada por $p(Y \mid \beta, X)$.

El modelo básico y clásico asume que la verosimilitud para las variables de interés es

$$Y \mid \theta, \sigma^2, X \sim Normal_n(X\beta, \sigma^2 I_n)$$

en donde I_n denota la matriz identidad de orden $n \times n$. Por supuesto, el modelo normal no es el único que se puede postular como verosimilitud para los datos.

Parámetros independientes

Suponiendo que los parámetros son independientes previa; es decir que la distribución previa conjunta está dada por

$$p(\beta, \sigma^2) = p(\beta)p(\sigma^2)$$

Como es natural, la distribución previa del vector de parámetros β es normal, aunque esta vez la matriz de varianzas no va a depender del otro parámetro σ^2, por lo tanto se tiene que

$$\beta \sim Normal_q(b, B)$$

Igualmente, el parámetro σ^2 no depende de β y es posible asignarle la siguiente distribución previa

$$\sigma^2 \sim IG\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$$

Distribución posterior

La distribución posterior conjunta de β y σ^2 puede ser escrita como

$$\begin{aligned}
p(\beta, \sigma^2 \mid Y, X) &\propto p(Y \mid \beta, \sigma^2) p(\beta) p(\sigma^2) \\
&\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}\left(Q(\beta) + S_e^2\right)\right\} \\
&\times \exp\left\{-\frac{1}{2}(\beta - b)' B^{-1} (\beta - b)\right\} (\sigma^2)^{-n_0/2-1} \exp\left\{-\frac{n_0 \sigma_0^2}{2\sigma^2}\right\} \\
&= (\sigma^2)^{-\frac{n+n_0}{2}-1} \exp\left\{-\frac{1}{2\sigma^2}\left[Q(\beta) + S_e^2 + n_0 \sigma_0^2\right]\right\} \\
&\times \exp\left\{-\frac{1}{2}(\beta - b)' B^{-1} (\beta - b)\right\}
\end{aligned} \tag{1}$$

Distribución posterior de β

La distribución posterior del parámetro β condicionado a σ^2, Y, X es

$$\beta \mid \sigma^2, Y, X \sim Normal_q(b_q, B_q)$$

donde

$$B_q = \left(B^{-1} + \frac{1}{\sigma^2} X'X\right)^{-1}$$
$$b_q = B_q \left(B^{-1} b + \frac{1}{\sigma^2} X'Y\right)$$

Distribución posterior de σ^2

La distribución posterior del parámetro σ^2 condicionado a β, Y, X es

$$\sigma^2 \mid \beta, Y, X \sim IG\left(\frac{n_1}{2}, \frac{n_1\sigma^2_\beta}{2}\right)$$

donde $n_1 = n + n_0$

$$\begin{aligned}
n_1\sigma^2_\beta = && Q(\beta) + S^2_e + n_0\sigma^2_0 \\
Q(\beta) = && (\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta}) \\
S^2_e = && (y - X\hat{\beta})'(y - X\hat{\beta})
\end{aligned}$$

y σ^2_0 es una estimación previa del parámetro de interés σ^2.

Modelo lineal en código de STAN

```
data {
  int<lower=0> n;   // Número de observaciones
  int<lower=0> K;   // Número de predictores
  matrix[n, K] x;   // Matrix de predictores
  vector[n] y;      // Vector respuesta
}
parameters {
  vector[K] beta;
  real<lower=0> sigma2;
}
model {
  to_vector(beta) ~ normal(0, 1000);
  sigma2 ~ inv_gamma(0.0001, 0.0001);
  y ~ normal(x * beta, sqrt(sigma2));  // likelihood
}
generated quantities {
    real ypred[n];  // vector de longitud n
    ypred = normal_rng(x * beta, sqrt(sigma2));
}
```

Proceso de estimación en `STAN`

Sea Y el logaritmo del ingreso medio por división administrativa del país.

Table 3: Logaritmo del ingreso

dam2	logingreso	luces_nocturnas	cubrimiento_cultivo	cubrimiento_urbano
05001	13.34	46.0570	2.0996	29.9636
05002	12.41	2.3771	1.3245	0.5746
05031	12.40	0.8686	0.1123	0.2894
05034	12.62	2.9262	1.5261	0.4212
05045	12.81	5.9329	0.5239	1.2359
05079	12.21	16.1735	2.5634	3.6992
05088	13.30	36.3252	8.6041	15.4245
05093	12.54	1.1088	2.4737	0.7746
05120	11.83	0.4323	1.8486	0.1544
05129	13.59	16.9501	0.9336	3.5665

Asociación de la variables

Table 4: Correlación con logaritmo del ingreso

Covariable	logingreso
luces_nocturnas	0.6312
cubrimiento_cultivo	0.2220
cubrimiento_urbano	0.4625
modificacion_humana	0.6765
accesibilidad_hospitales	-0.3857
accesibilidad_hosp_caminado	-0.4175

Preparando el código de STAN

Organizando datos para STAN

```
fitLm2 <-"www/00_Intro_bayes/Modelo/03_ModeloLm.stan"

Xdat <- model.matrix(
  logingreso ~ luces_nocturnas +
    cubrimiento_cultivo +  cubrimiento_urbano +
    modificacion_humana ,data = datalm)

sample_data <- list(n = nrow(datalm),
                    K = ncol(Xdat),
                    x = as.matrix(Xdat),
                    y = datalm$logingreso)
```

Preparando datos para el código de STAN

```r
model_fitLm2 <-
  stan(
    file = fitLm2,
    data = sample_data,
    warmup = 500,
    iter = 1000,
    verbose = FALSE,
    cores = 4
  )

saveRDS(model_fitLm2,
           "www/00_Intro_bayes/Modelo/model_fitLm2.rds")
```

Resultados de la estimación del parámetro β y σ^2 es:

```r
tabla_coef <- summary(model_fitLm2,
        pars = c("beta", "sigma2"))$summary
```

	mean	se_mean	sd	2.5%	25%	50%	75%	97.5%	n_eff	Rhat
beta[1]	11.9030	0.0030	0.0757	11.7515	11.8538	11.9020	11.9559	12.0454	645.7	1.0074
beta[2]	0.0129	0.0001	0.0040	0.0054	0.0102	0.0129	0.0157	0.0209	851.1	1.0028
beta[3]	-0.0008	0.0000	0.0015	-0.0037	-0.0017	-0.0008	0.0002	0.0021	2522.2	1.0010
beta[4]	-0.0097	0.0001	0.0051	-0.0197	-0.0130	-0.0097	-0.0061	0.0004	1245.5	0.9992
beta[5]	1.6917	0.0104	0.2630	1.1960	1.5089	1.6849	1.8688	2.2178	638.3	1.0075
sigma2	0.1032	0.0002	0.0069	0.0909	0.0981	0.1028	0.1075	0.1178	964.7	1.0056

Convergencias de las cadenas el parámetro θ

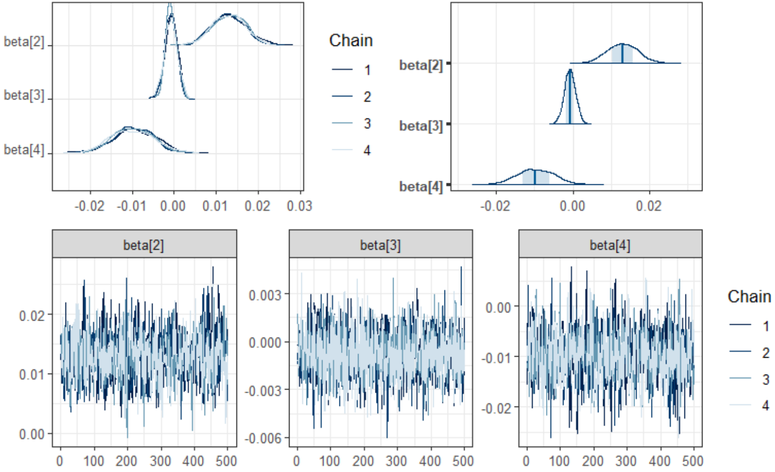


Figure 7: Cadenas para beta

Convergencias de las cadenas el parámetro σ^2

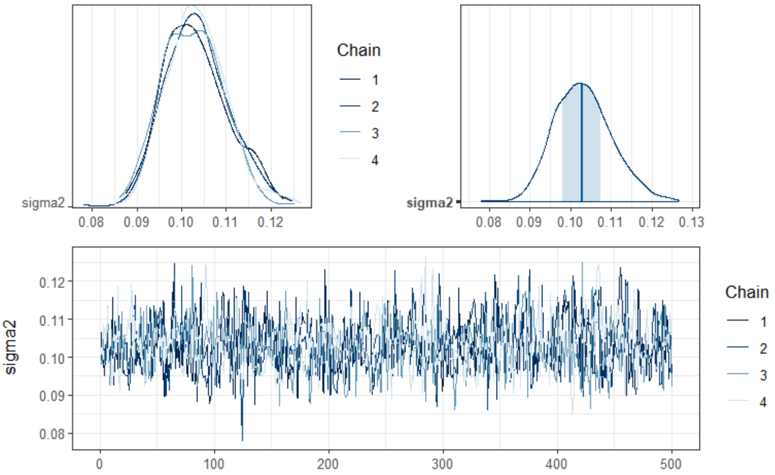


Figure 8: Cadenas para sigma2

Chequeo predictivo posterior del ingreso

```
y_pred_B <- as.array(model_fitLm2, pars = "ypred") %>%
  as_draws_matrix()

rowsrandom <- sample(nrow(y_pred_B), 100)

log_pred2 <- y_pred_B[rowsrandom,]
y_pred2 <- exp(log_pred2)-1

ppc_dens_overlay(
  y = datalm$logingreso, log_pred2)/
ppc_dens_overlay(
  y = as.numeric(exp(datalm$logingreso) - 1), y_pred2) +
  xlim(0, 800000)
```

Chequeo predictivo posterior del ingreso

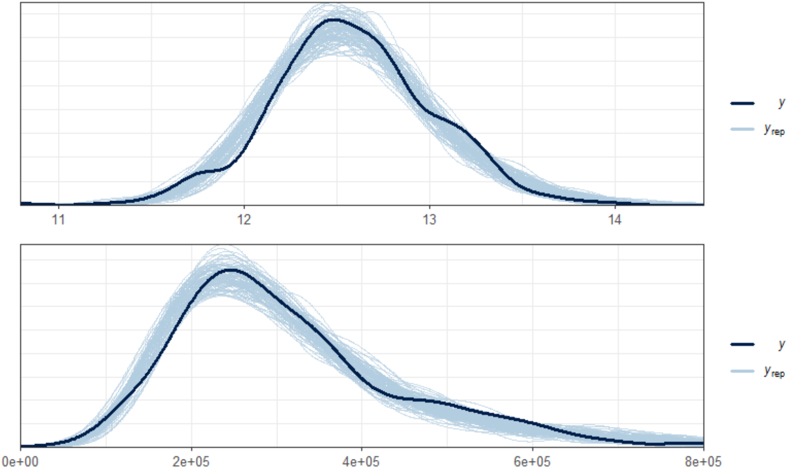


Figure 9: PPC para el ingreso