Edit or run this notebook

Nonparametric probability estimation

Statistics III – Dr. Arturo Erdely

The random variables $Z_i := \mathbb{1}_{\{X_i \in B\}}$ for $i \in \{1, \ldots, n\}$ are i.i.d Bernoulli with unknown parameter $\theta = \mathbb{P}(X \in B)$. Using a $\text{Uniform}(0, 1)$ non-informative prior distribution for θ the posterior distribution is $\text{Beta}(1 + nT_n(B), 1 + n(1 - T_n(B)))$ where $Tn(B) = \frac{1}{n} \sum_{i=1}^n Z_i$ y a nonparametric estimation of $\mathbb{P}(X \in B)$. Under a quadratic loss penalization, the bayesian point estimate for $\theta = \mathbb{P}(X \in B)$ is the posterior *expected value* which is the following:

$$\theta^* = \frac{1 + nT_n(B)}{2 + n}$$

An interval estimation for $\theta = \mathbb{P}(X \in B)$ with probability $0 < \gamma < 1$ is an interval $[\theta_1, \theta_2]$ such that $F_\theta(\theta_2) - F_\theta(\theta_1) = \gamma$, where F_θ is the posterior distribution of θ, and such that the interval length $\theta_2 - \theta_1$ is minimum. Since $\theta_2 = F_\theta^{-1}(\gamma + F_\theta(\theta_1))$, where $0 \leq \theta_1 \leq F_\theta^{-1}(1 - \gamma)$, the minimum length interval must be the solution to minimize the following function:

$$h(z) = F_\theta^{-1}(\gamma + F_\theta(z)) - z \,, \qquad 0 \leq z \leq F_\theta^{-1}(1 - \gamma)$$

```
using Distributions
```

```
Tn (generic function with 1 method)
```

```
EDA (generic function with 1 method)
```

```
Bn (generic function with 2 methods)
```

Edit or run this notebook

```julia
function Bn(interval::String, obs, γ = 0.95)
    # using: Distributions
    # Dependencies: Tn, EDA
    n = length(obs)
    tn = Tn(interval, obs)
    α, β = 1 + n*tn, 1 + n*(1 - tn) # posterior parameters
    Θ = Beta(α, β) # posterior distribution
    θmedia, θmediana = mean(Θ), median(Θ)
    h(z) = (quantile(Θ, γ + cdf(Θ, z[1])) - z[1]) * Inf^(z[1] > quantile(Θ, 1 - γ))
    sol = EDA(h, [0], [quantile(Θ, 1 - γ)])
    θ₁ = sol[1][1]
    θ₂ = quantile(Θ, γ + cdf(Θ, θ₁))
    estimación = (insesgado = tn, media = θmedia, mediana = θmediana, intervalo =
(θ₁, θ₂))
    return estimación
end
```

Let W be a $\text{Normal}(\mu = -2, \sigma = 3)$ random variable. Then $\mathbb{P}(0 < W < 3) = F_W(3) - F_W(0)$, that is:

```
0.20470218527410822
```

```julia
begin
    W = Normal(-2, 3)
    P = cdf(W, 3) - cdf(W, 0)
end
```

Let's now simulate $1,000$ observations from W and make like we forget its distribution, and calculate a nonparametric point and interval estimations for $\mathbb{P}(0 < W < 3)$.

```
▸ (insesgado = 0.199, media = 0.199601, mediana = 0.199401, intervalo = (0.175055, 0.224483
```

```julia
begin
    wobs = rand(W, 1_000)
    estimP = Bn("]0,3[", wobs)
end
```