Bayesian SPC for Count Data

Panagiotis Tsiamyrtzis, Dept. of Statistics, AUEB, Greece

Douglas M. Hawkins, School of Statistics, U of M, USA

Introduction - Frequentist approach

- For count data the standard SPC control chart method is the u-Control Chart (u-CC), where the quality characteristic is defined as the number of defects per inspection unit (assumed to follow a Poisson distribution).

Introduction - Frequentist approach

- For count data the standard SPC control chart method is the u-Control Chart (u-CC), where the quality characteristic is defined as the number of defects per inspection unit (assumed to follow a Poisson distribution).
- Construction of the u-CC will require the collection of Phase-I data, which are considered to be iid observations. These data will then be used to estimate the parameter of interest being In Control and test the upcoming observations (Phase-II data).

The Problems

- The u-CC will suffer if:

The Problems

- The u-CC will suffer if:
  - parameter of interest drifts during Phase-I

The Problems

- The u-CC will suffer if:
  - parameter of interest drifts during Phase-I
  - the data are over/under dispersed

The Problems

- The u-CC will suffer if:
  - parameter of interest drifts during Phase-I
  - the data are over/under dispersed
  - serial dependence exits

The Problems

- The u-CC will suffer if:
  - parameter of interest drifts during Phase-I
  - the data are over/under dispersed
  - serial dependence exits

- Even if none of the above exists we are not able to:

The Problems

- The u-CC will suffer if:
  - parameter of interest drifts during Phase-I
  - the data are over/under dispersed
  - serial dependence exits

- Even if none of the above exists we are not able to:
  - do formal testing during Phase-I/short runs

The Problems

- The u-CC will suffer if:
  - parameter of interest drifts during Phase-I
  - the data are over/under dispersed
  - serial dependence exits

- Even if none of the above exists we are not able to:
  - do formal testing during Phase-I/short runs
  - use available prior information regarding the parameter

Introduction - Bayesian approach

- **Hoadley's (1981) Quality Measurement Plan (QMP)**

$$X_n|\theta_n \sim Poisson(e_n\theta_n)$$

where the $E[X_n]$ involves an unknown parameter θ_n multiplying a known expectation coefficient e_n.

Introduction - Bayesian approach

- **Hoadley's (1981) Quality Measurement Plan (QMP)**

$$X_n|\theta_n \sim Poisson(e_n\theta_n)$$

where the $E[X_n]$ involves an unknown parameter θ_n multiplying a known expectation coefficient e_n.

The process parameter θ_n is a random variable:

$$\theta_n|(\alpha,\beta)\sim Gamma(\alpha,\beta)$$

Introduction - Bayesian approach

- **Hoadley's (1981) Quality Measurement Plan (QMP)**

$$X_n|\theta_n \sim Poisson(e_n\theta_n)$$

where the $E[X_n]$ involves an unknown parameter θ_n multiplying a known expectation coefficient e_n.

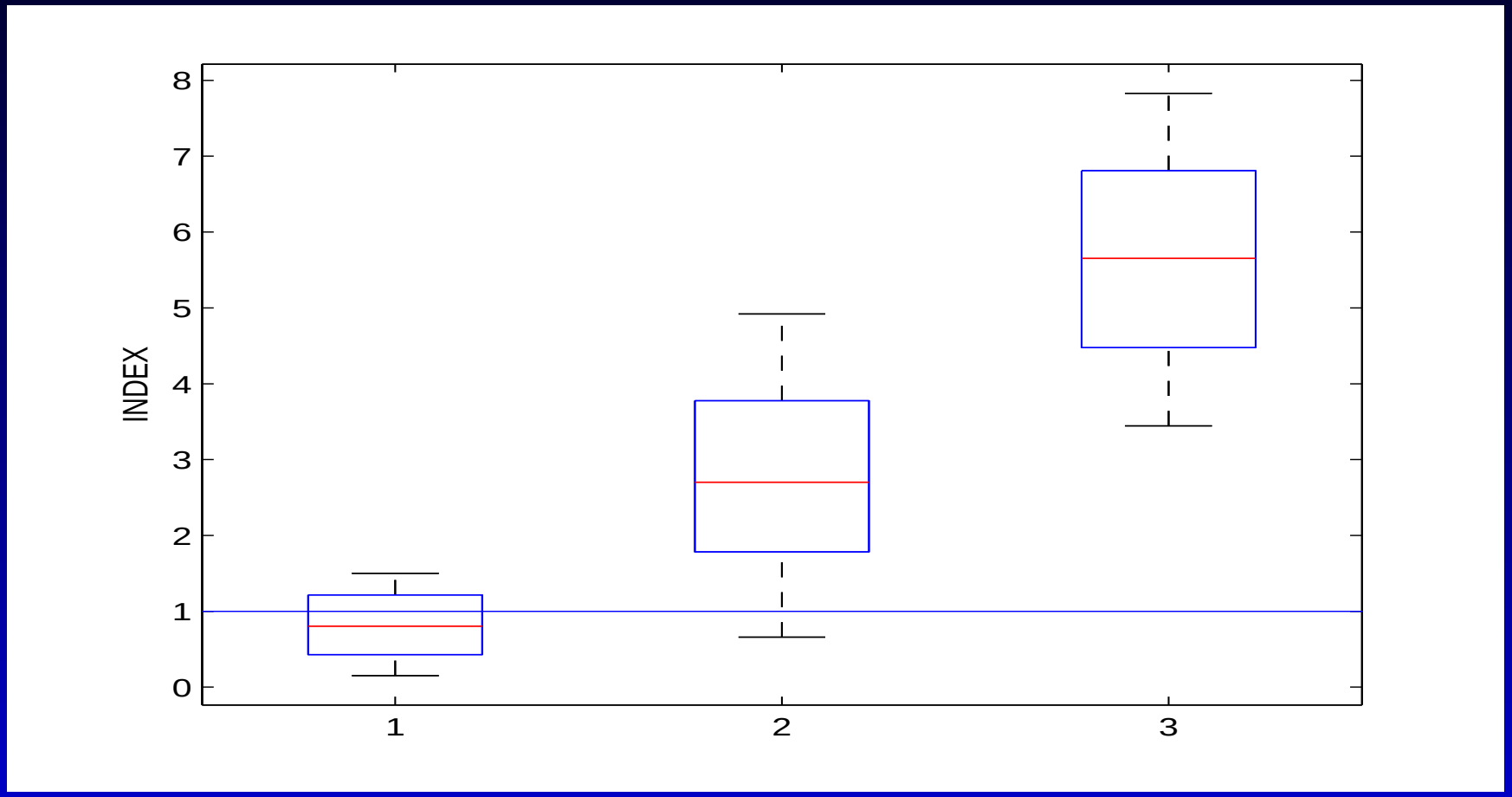
The process parameter θ_n is a random variable:

$$\theta_n|(\alpha,\beta)\sim Gamma(\alpha,\beta)$$

leading to the posterior distribution

$$\theta_n|(x_n,\alpha,\beta) \sim Gamma(\alpha + x_n, \beta + e_n)$$

Graphical representation of QMP



A QMP plot of a *Normal*, an *Alert* and a *Below Normal* scenarios

Our Approach

At each stage n of the process we will have some prior information for the parameter of interest (θ_n) which will be modeled using a change point model.

Our Approach

At each stage n of the process we will have some prior information for the parameter of interest (θ_n) which will be modeled using a change point model.

When the observation X_n will become available Bayes theorem will provide the posterior distribution which will be used to decide whether we are still *In Control* or not.

Our Approach

At each stage n of the process we will have some prior information for the parameter of interest (θ_n) which will be modeled using a change point model.

When the observation X_n will become available Bayes theorem will provide the posterior distribution which will be used to decide whether we are still *In Control* or not.

When the action is to leave the process operating, the posterior of this stage will be used as a prior for the next stage (sequentially updated).

Statistical Modeling

At time 0, prior to any process reading we have some prior distribution

$$\theta_0 \sim Gamma(\alpha_0, \beta_0)$$

Statistical Modeling

At time 0, prior to any process reading we have some prior distribution

$$\theta_0 \sim Gamma(\alpha_0, \beta_0)$$

Moving between successive stages we model the parameter using a change point model:

$$\theta_n | \theta_{n-1} \sim \left\{ \begin{array}{lll} \theta_{n-1} & \text{with prob. } 1 - p_1 - p_2 \\ \lambda_1 \theta_{n-1} & \text{with prob. } \quad p_1 \\ \lambda_2 \theta_{n-1} & \text{with prob. } \quad p_2 \end{array} \right\}$$

where $0 < \lambda_1 < 1$ (quality improvement), $\lambda_2 > 1$ (quality degradation)

Statistical Modeling

At time 0, prior to any process reading we have some prior distribution

$$\theta_0 \sim Gamma(\alpha_0, \beta_0)$$

Moving between successive stages we model the parameter using a change point model:

$$\theta_n|\theta_{n-1} \sim \left\{ \begin{array}{lll} \theta_{n-1} & \text{with prob. } 1-p_1-p_2 \\ \lambda_1\theta_{n-1} & \text{with prob.} \quad p_1 \\ \lambda_2\theta_{n-1} & \text{with prob.} \quad p_2 \end{array} \right\}$$

where $0 < \lambda_1 < 1$ (quality improvement), $\lambda_2 > 1$ (quality degradation)

Regarding the observed data we will have:

$$X_n|\theta_n \sim Poisson(\theta_n)$$

Model Evolution - Approximation

- At each stage n of the process the posterior distribution of $\theta_n|\mathbf{X}_n$ will be a mixture of 3^n Gamma components:

$$p(\theta_n|\mathbf{X}_n) \sim \sum_{j=0}^{3^n-1} w_j^{(n)} Gamma\left(\alpha^{(n)}, \ \ \beta_j^{(n)}\right)$$

where the weights, and parameters are given recursively.

Model Evolution - Approximation

- At each stage n of the process the posterior distribution of $\theta_n|\mathbf{X}_n$ will be a mixture of 3^n Gamma components:

$$p(\theta_n|\mathbf{X}_n) \sim \sum_{j=0}^{3^n-1} w_j^{(n)} Gamma\left(\alpha^{(n)},\ \ \beta_j^{(n)}\right)$$

  where the weights, and parameters are given recursively.

- As n increases we can approximate the posterior mixture with another mixture having far fewer components. This could be done by pruning all the components which have 'negligible' weights in the mixture since all their descendants will have even smaller weights representing the most unlikely scenarios of the model evolution

Decision Making

Once the posterior distribution of $\theta_n|\mathbf{X}_n$ becomes available we can easily obtain point/interval estimates of θ_n.

Decision Making

Once the posterior distribution of $\theta_n|\mathbf{X}_n$ becomes available we can easily obtain point/interval estimates of θ_n.

Usually we are interested in the sequence of Hypothesis Testing:

$$\left\{ \begin{array}{ll} H_0: & \theta_n \leq M \\ H_1: & \theta_n > M \end{array} \right\}$$

We can calculate $p_n = P(\theta_n < M|\mathbf{X}_n)$ and decide to reject H_0 if $p_n < c^*$. Alternatively we could employ Bayes Factor.

A time series plot of the p_n probabilities will offer further insight in the behavior of the parameter of interest.

Forecasting

Based on the observations $\mathbf{X}_n = \{X_1, X_2, \ldots, X_n\}$ we can derive the predictive distribution of the next (unseen) observable:

$$P(X_{n+1}|\mathbf{X}_n) = \int f(X_{n+1}|\theta_{n+1})\pi(\theta_{n+1}|\mathbf{X}_n)d\theta_{n+1}$$

which will be a mixture of 3^{n+1} Poisson - Gamma distributions.

The predictive distribution can be used not only for predictions but for model assessment purposes as well.

Example

In an assembly line of complex electrical equipment we count the number of defect per unit. We have measurements for 12 successive days (Hansen and Ghare, 1987):

t	1	2	3	4	5	6	7	8	9	10	11	12
x_t	3.86	5	4.71	3	4	4.14	5.17	4.88	4.83	10	3.88	2.33

For the initial prior distribution we used:

$$\alpha_0 = 4, \qquad \beta_0 = 1$$

while for the rest of the nuisance parameters we chose:

$$\lambda_1 = 0.5, \qquad \lambda_2 = 1.5, \qquad p_1 = p_2 = 0.2$$

Example cont.

The tolerable upper threshold value of the expected defects per unit was $M = 6$. Applying the exact model to the data, gives the posterior probabilities $p_n = P(\theta_n < M|\mathbf{X}_n)$ at the end of each time period n:

t	1	2	3	4	5	6	7	8	9	10	11	12
x_t	3.86	5	4.71	3	4	4.14	5.17	4.88	4.83	10	3.88	2.33
p_n	.91	.85	.83	.93	.92	.91	.84	.81	.79	.24	.65	.93

Under the $0-1$ loss function we would have reject the null hypothesis at stage 10 of the process. However it is clear that right after this stage the process gets down to acceptable levels, indicating observation 10 is more likely an outlier as opposed to the beginning of a persistent trend leading to quality degradation

Extending the Model

Instead of using estimates of the nuisance parameters based on ‘similar’ type processers we could adopt a purely Bayesian approach where we use a prior distribution on the vector of nuisance parameters.

Then a joint posterior will be obtained and integrating out the nuisance parameters will provide the distribution of the parameter of interest

Thank you!