Journal of Data Science 0 (), 1–8 DOI:
0000

Supplementary

S.1 Pólya-Gamma augmentation for our model

For completeness, we will first introduce the data augmentation method with some useful results and then we shall write down the joint distribution of the augmented model.

For $b > 0$ and $c \in \Re$, a positive random variable w follows Pólya Gamma distribution $\mathrm{PG}(b, c)$ if

$$w \sim \mathrm{PG}(b,c) \quad \overset{d}{=} \frac{1}{2\pi^2}\sum_{k=1}^{\infty}\frac{g_k}{(k-\frac{1}{2})^2+\frac{c^2}{4\pi^2}} \tag{S1}$$

where the g_k's are independently distributed according to Gamma(b,1). We will apply the following two important properties involving Pólya-Gamma distribution in our setting. Firstly, let us denote the density function of $PG(b, c)$ to be $p(w; b, c)$. For $u \in \Re$ and $\eta \in \Re$,

$$\frac{(e^{\eta})^u}{(1+e^{\eta})^b} = 2^{-b}e^{(u-b/2)\eta}\int_0^{\infty} e^{-\frac{w\eta^2}{2}}p(w;b,0)dw. \tag{S2}$$

The first property suggests how the augmentation helps in replacing the difficult logistic likelihood with a easier Normal likelihood. The second property gives a relationship between Pólya Gamma distributions,

$$p(w;b,c) \propto e^{-\frac{c^2 w}{2}}p(w;b,0). \tag{S3}$$

Let us first write down the joint distribution of our model before data augmentation. Denote $\mathscr{D}$ to be all the observed data and Θ be the model parameters. So the joint density is of the form:

$$\begin{aligned} f(\mathscr{D};\Theta) \quad = \quad & \prod_{i=1}^{I}\frac{\Gamma(y_i+r)}{\Gamma(r)y_i!}\frac{\left[\exp(\sum_{d=1}^{D}\prod_{k=1}^{3}a_{i_k,d}^{(k)})\right]^{y_i}}{\left[1+exp(\sum_{d=1}^{D}\prod_{k=1}^{3}a_{i_k,d}^{(k)})\right]^{y_i+r}} \times \prod_{i=1}^{I}\prod_{l=1}^{L_i}\frac{\left[exp(\xi\phi_i+\beta^T\mathbf{z}_{il})\right]^{x_{il}}}{\left[1+exp(\xi\phi_i+\beta^T\mathbf{z}_{il})\right]} \\ \times & \prod_{i=1}^{I}\frac{\tau_{\phi}^{1/2}}{\sqrt{2\pi}}\exp\left(-\frac{\tau_{\phi}}{2}\left(\phi_i-\sum_{d=1}^{D}\prod_{k=1}^{3}a_{i_k,d}^{(k)}\right)^2\right) \times \prod_{d=1}^{D}\prod_{k=1}^{3}\prod_{j=1}^{I_k}(2\pi)^{-1/2}\lambda_d^{1/2}\exp\left(-\frac{\lambda_d}{2}a_{jd}^{(k)2}\right) \\ \times & \prod_{d=1}^{D}\frac{\epsilon^{\epsilon}}{\Gamma(\epsilon)}\lambda_d^{\epsilon-1}e^{-\epsilon\lambda_d} \times \frac{\delta^{\delta}}{\Gamma(\delta)}\tau_{\phi}^{\delta-1}e^{-\delta\tau_{\phi}} \times \frac{\tau_{\xi}^{1/2}}{\sqrt{2\pi}}\exp\left(-\frac{\tau_{\xi}}{2}\xi^2\right) \times \frac{|\Omega|^{1/2}}{(2\pi)^{r/2}}\exp\left(-\frac{1}{2}\beta^T\Omega\beta\right). \end{aligned} \tag{S4}$$

The first line includes the two logistic likelihood. Therefore we shall augment them with two sets of Pólya-Gamma random variables, namely $\{w_i\}_{i\in\{1,\ldots,I\}}$ and $\{v_{il}\}_{l\in\{1,\ldots,L_i\},i\in\{1,\ldots,I\}}$ with w_i following $PG(y_i + r; 0)$ and v_{ij} following $PG(1; 0)$. Denote $\mathscr{U}$ to be the set of all auxiliary Pólya-Gamma random variables, the complete joint distribution including the auxiliary variales becomes

$$\begin{aligned} f(\mathscr{D};\mathscr{U},\Theta) &= \prod_{i=1}^{I}\frac{\Gamma(y_i+r)}{\Gamma(r)y_i!2^{y_i+r}}\exp\left(\frac{y_i-r}{2}\sum_{d=1}^{D}\prod_{k=1}^{3}a_{i_k,d}^{(k)}\right)\exp\left(-\frac{w_i\left(\sum_{d=1}^{D}\prod_{k=1}^{3}a_{i_k,d}^{(k)}\right)^2}{2}\right)p(w_i;y_i+r,0)\\ &\times \prod_{i=1}^{I}\prod_{l=1}^{L_i}2^{-1}\exp\left((x_{il}-1/2)\left(\xi\phi_i+\beta^T\mathbf{z}_{il}\right)\right)\exp\left(-\frac{v_{il}\left(\xi\phi_i+\beta^T\mathbf{z}_{il}\right)^2}{2}\right)p(v_{il};1,0)\\ &\times \prod_{i=1}^{I}\frac{\tau_\phi^{1/2}}{\sqrt{2\pi}}\exp\left(-\frac{\tau_\phi}{2}\left(\phi_i-\sum_{d=1}^{D}\prod_{k=1}^{3}a_{i_k,d}^{(k)}\right)^2\right)\times\prod_{d=1}^{D}\prod_{k=1}^{3}\prod_{j=1}^{I_k}(2\pi)^{-1/2}\lambda_d^{1/2}\exp\left(-\frac{\lambda_d}{2}a_{jd}^{(k)2}\right)\\ &\times \prod_{d=1}^{D}\frac{\epsilon^\epsilon}{\Gamma(\epsilon)}\lambda_d^{\epsilon-1}e^{-\epsilon\lambda_d}\times\frac{\delta^\delta}{\Gamma(\delta)}\tau_\phi^{\delta-1}e^{-\delta\tau_\phi}\times\frac{\tau_\xi^{1/2}}{\sqrt{2\pi}}\exp\left(-\frac{\tau_\xi}{2}\xi^2\right)\times\frac{|\Omega|^{1/2}}{(2\pi)^{r/2}}\exp\left(-\frac{1}{2}\beta^T\Omega\beta\right). \quad \text{(S5)} \end{aligned}$$

S.2 Variational EM steps

Below we write down the variational distributions of each component. The expectations E_q refers to the most updated variational distributions of the variables.

- Update q_j for $\Theta_j = a_{ld}^{(k)}$ for $d \in \{1, ..., D\}$, $k \in \{1, 2, 3\}$ and $l \in \{1, ..., I_k\}$,

$$q_j(a_{ld}^{(k)}) = TN_{(0,\infty)}\left(\mu_{ldk}^a, (\omega_{ldk}^a)^{-1}\right) \quad \text{(S6)}$$

  where

$$\omega_{ldk}^a = \sum_{i:i_k=l}\left(\left(E_q(w_i)+E_q(\tau_\phi)\right)\prod_{k'\neq k}E_q\left(a_{i_{k'}d}^{(k')2}\right)\right)+E_q(\lambda_d),$$

$$\mu_{ldk}^a = \frac{\sum_{i_k=l}\left(\prod_{k'\neq k}E_q\left(a_{i_{k'}d}^{(k')}\right)\left(\frac{y_i-r}{2}-E_q(w_i)\sum_{d'\neq d}\prod_{k=1}^{3}E_q(a_{i_kd'}^{(k)})+E_q(\tau_\phi)\left(E(\phi_i)-\sum_{d'\neq d}\prod_{k=1}^{3}E_q(a_{i_kd'}^{(k)})\right)\right)\right)}{\omega_{ldk}^a}.$$

- Update q_j for $\Theta_j = \lambda_d$ for $d \in \{1, ..., D\}$,

$$q_j(\lambda_d) = Gamma\left(\sum_{k=1}^{3}I_k/2+\epsilon, \sum_{k=1}^{3}\sum_{l=1}^{I_k}E_q(a_{ld}^{(k)})/2+\epsilon\right). \quad \text{(S7)}$$

- Update q_j for $\Theta_j = \phi_i$ for $i \in \{1, ..., I\}$

$$q_j(\phi_i) = N\left(\mu_{\phi_i}, (\omega_{\phi_i})^{-1}\right) \quad \text{(S8)}$$

  where

$$\omega_{\phi_i} = E_q(\tau_\phi)+E_q(\xi^2)\sum_{l=1}^{y_i}E_q(v_{il}),$$

$$\mu_{\phi_i} = \frac{E_q(\tau_\phi)\sum_{d=1}^{D}\prod_{k=1}^{3}E_q(a_{i_kd}^{(k)})+E_q(\xi)\left(\sum_{l=1}^{y_i}(x_{il}-0.5)-\sum_{l=1}^{y_i}v_{il}\mathbf{z}_{il}^TE_q\beta\right)}{\omega_{\phi_i}}.$$

- Update q_j for $\Theta_j = w_i$ for $i \in \{1, ..., I\}$

$$q_j(w_i) = PG\left(y_i + r, \sqrt{E_q\left(\sum_{d=1}^{D}\prod_{k=1}^{3} a_{i_k d}^{(k)}\right)^2}\right). \tag{S9}$$

- Update q_j for $\Theta_j = v_{il}$ for $l \in \{1, ..., J_i\}, i \in \{1, ..., I\}$

$$q_j(v_{il}) = PG\left(1, \sqrt{E_q\left(\xi\phi_i + \mathbf{z}_{il}^T\beta\right)^2}\right). \tag{S10}$$

- Update q_j for $\Theta_j = \tau_\phi$,

$$q_j(\tau_\phi) = Gamma\left(\frac{\sum_{k=1}^{3} I_k}{2} + \delta, \frac{\sum_{i=1}^{I} E_q\left(\phi_i - \sum_{d=1}^{D}\prod_{k=1}^{3} a_{i_k d}^{(k)}\right)^2}{2} + \delta\right). \tag{S11}$$

- Update q_j for $\Theta_j = \xi$,

$$q_j(\xi) = N\left(\mu_\xi, (\omega_\xi)^{-1}\right) \tag{S12}$$

where

$$\omega_\xi = \tau_\xi + \sum_{i=1}^{I}\left(E_q(\phi_i^2)\sum_{l=1}^{y_i} E_q(v_{il})\right),$$

$$\mu_\xi = \frac{\sum_{i=1}^{I}\left(E_q(\phi_i)\sum_{l=1}^{J_i}\left(x_{il} - 0.5 - E_q(v_{il}\mathbf{z}_{il}^T\beta)\right)\right)}{\omega_\xi}.$$

- Update q_j for $\Theta_j = \beta$,

$$q_j(\beta) = N_m\left(\mu_\beta, (\Omega_\beta)^{-1}\right) \tag{S13}$$

where

$$\Omega_\beta = \sum_{i=1}^{I}\sum_{l=1}^{y_i} E_q(v_{il})\mathbf{z}_{il}\mathbf{z}_{il}^T + \Omega,$$

$$\mu_\beta = \Omega_\beta^{-1}\left(\sum_{i=1}^{I}\sum_{l=1}^{y_i}(x_{il} - 0.5)\mathbf{z}_{il} - E_q(\xi)\sum_{i=1}^{I}\sum_{l=1}^{y_i} E_q(\phi_i)E_q(v_{il})\mathbf{z}_{il}\right).$$

After updating the variational distributions, we search r that maximize the ELBO. In this step we only have to maximize the part of the ELBO that is affected by r, which is the following function $g(r)$:

$$g(r) = \sum_{i=1}^{I}\log\Gamma(y_i + r) - I\log\Gamma(r) - K_g r - \sum_{i=1}^{I} KL\left(\frac{PG(\omega; y_i + \tilde{r}, w_c)}{PG(\omega; y_i + r, w_c)}\right) \tag{S14}$$

where $\tilde{r}$ is the current estimate of r, and the constant w_c and K_g are,

$$w_c = \sqrt{E_q\left(\sum_{d=1}^{D}\prod_{k=1}^{3} a_{i_k d}^{(k)}\right)^2}$$

and

$$K_g = I\log 2 + \frac{1}{2}\sum_{i=1}^{I}\sum_{d=1}^{D}\prod_{k=1}^{3} E_q(a_{i_k d}^{(k)}) + \sum_{i=1}^{I}\log\cosh\left(\frac{1}{2}w_c\right)$$

respectively. The last sum in (S14) is difficult to calculate as there is no closed form for the KL divergence. Thus we follow Soulat et al. (2021) to use moment-matched Gamma distributions to approximate the Pólya-Gamma distributions to calculate the KL divergence values. Applying this method, $g(r)$ in (S14) can be approximated by,

$$\sum_{i=1}^{I}\log\Gamma(y_i+r) - I\log\Gamma(r) - K_g r + \sum_{i=1}^{I}\left[(y_i+r)C_g h((y_i+\tilde{r})C_g) - \log\Gamma((y_i+r)C_g)\right],$$

where h denotes the digamma function and

$$C_g = \frac{w_c Cosh^2(w_c/2)tanh^2(w_c/2)}{sinh(w_c) - w_c}.$$

After dropping terms with no r and rearranging the terms, eventually the function to be optimized is,

$$\tilde{g}(r) \quad = \quad \sum_{i=1}^{I}\left[\log\Gamma(y_i+r) - \log\Gamma((y_i+r)C_g) - \log\Gamma(r)\right] + \left(\sum_{i=1}^{I} C_g h((y_i+\tilde{r})C_g) - K_g\right) r. \tag{S15}$$

To study the shape of $\tilde{g}(r)$, we differentiate it twice to get

$$\tilde{g}^{''}(r) = \sum_{i=1}^{I}\left[T(y_i+r) - T(r) - T\left(C_g(y_i+r)\right)C_g^2\right],$$

where T denotes the trigamma function. As we know that trigamma function is positive and decreasing on the positive axis, the term $T(y_i+r) - T(r)$ is negative for positive r. In addition, the constant C_g is obviously positive. Thus we see that $\tilde{g}^{''}(r) < 0$ and hence $\tilde{g}(r)$ is strictly concave for positive r. Hence the maximum can be easily located numerically.

S.3 More graphs for the analysis in section 4

Below are the graphs of the seasons not reported in Figure 2:

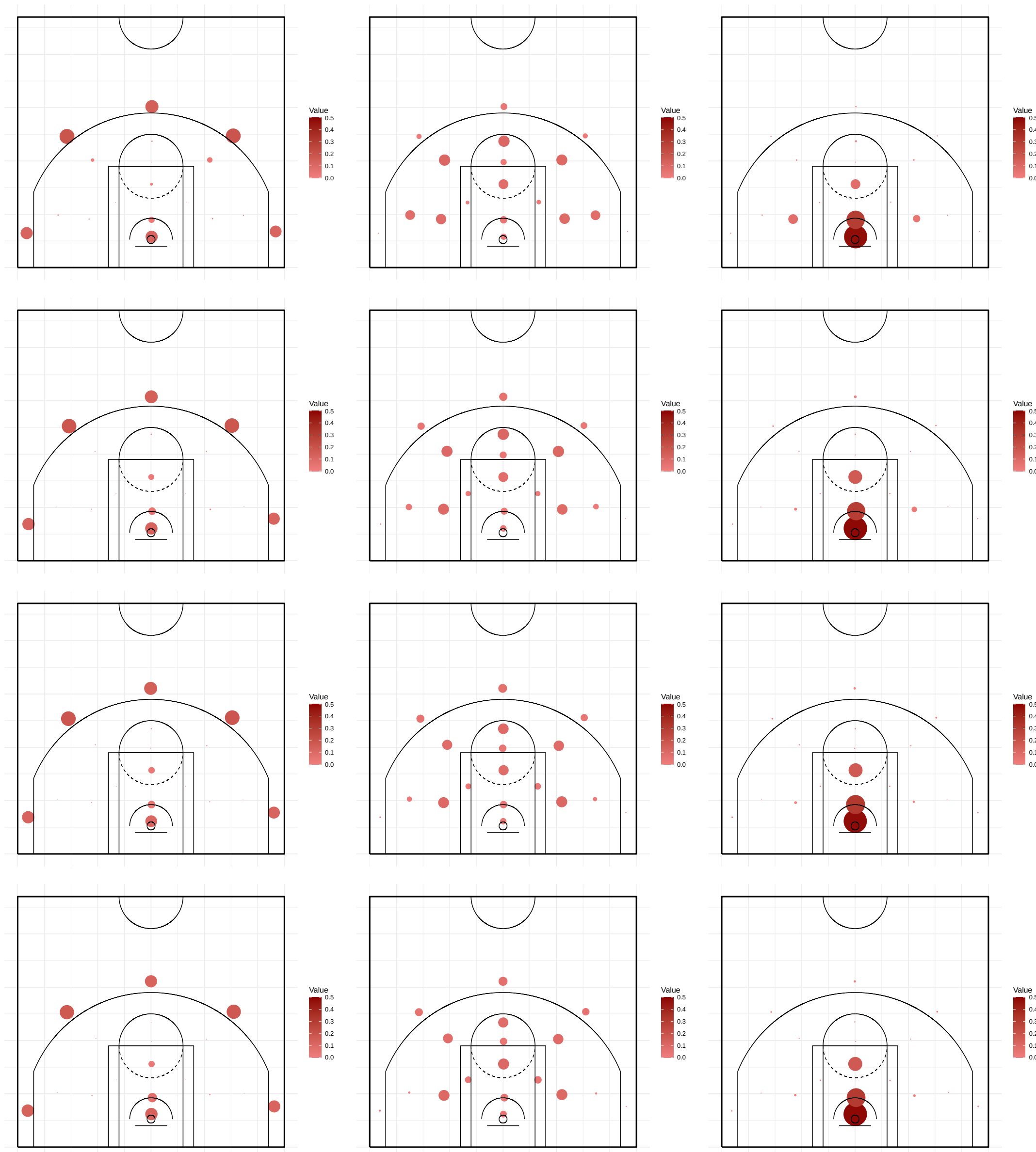


Figure 1: The first, second, third and fourth rows are for seasons 2016-17, 2018-19, 2020-21 and 2022-23 respectively. The columns correspond to the three latent factors of the location mode. Zones with redder colors and larger areas are with larger posterior means for the latent factor.

Below are the graphs of the seasons not reported in Figure 3:

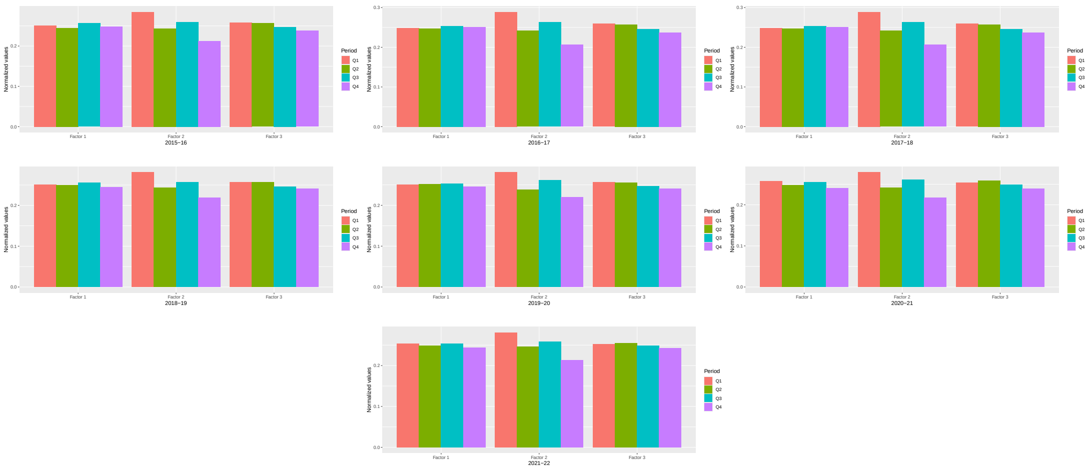
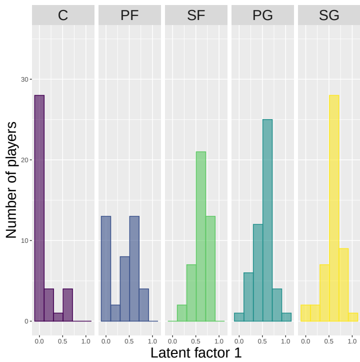
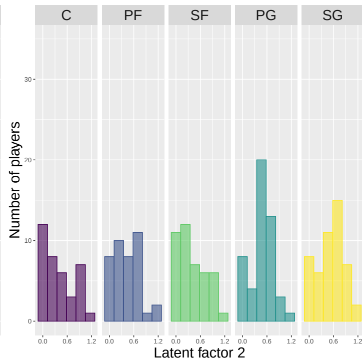


Figure 2: The three bar graphs are correspond to the three latent factors of the period mode of the regular seasons 2015-16 to 2021-22. Each bar graph shows the normalized posterior means of the factor values for the four quarters.

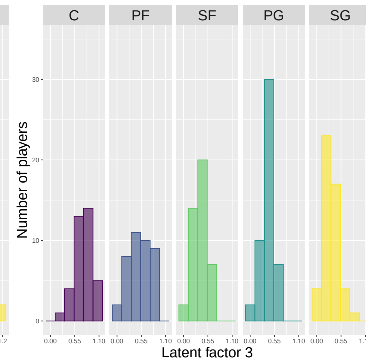
Below are the graphs of the seasons not reported in Figure 4:



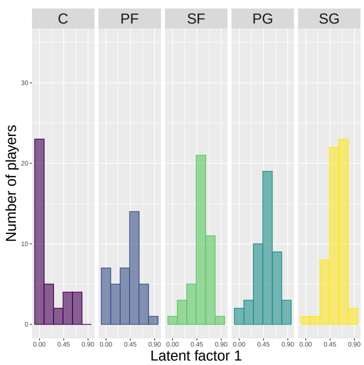
2015-16



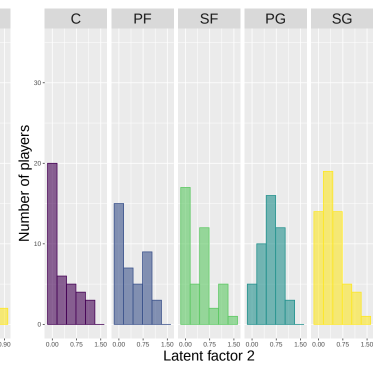
2015-16



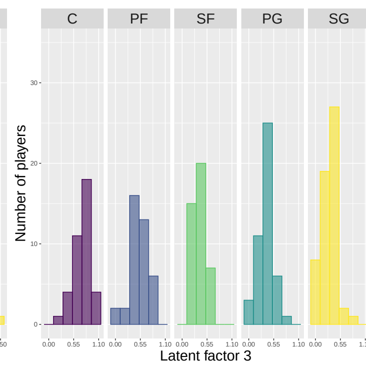
2015-16



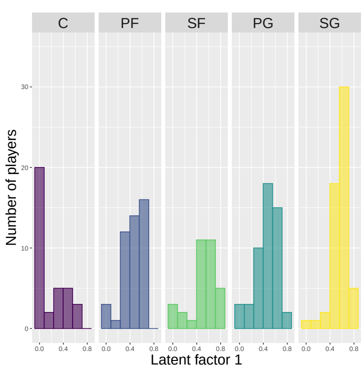
2016-17



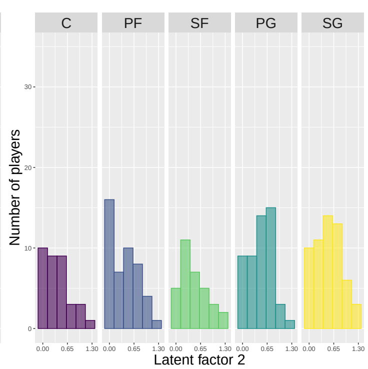
2016-17



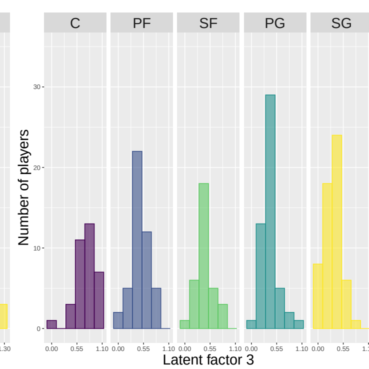
2016-17



2017-18



2017-18



2017-18

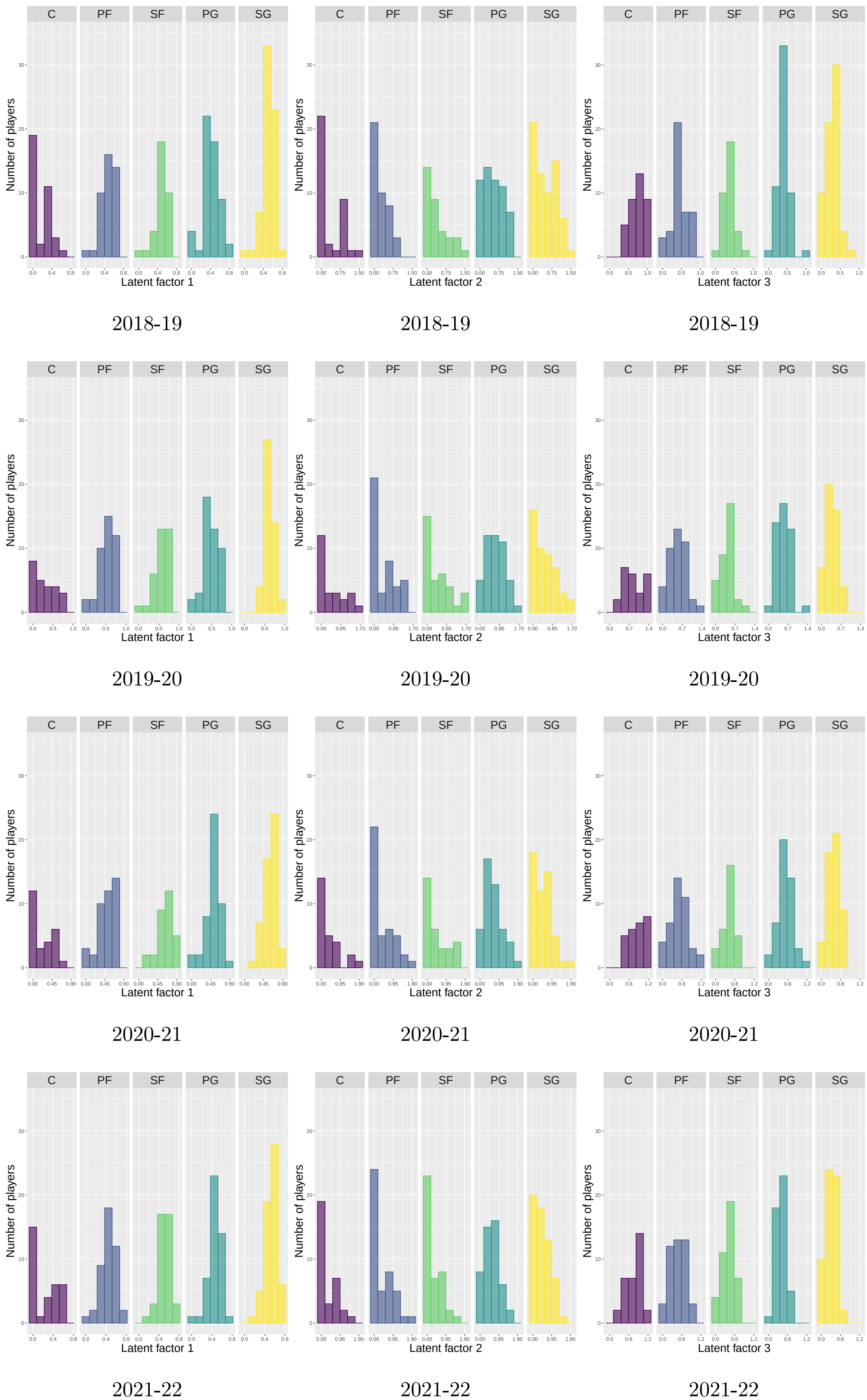


Figure 3: Each panel demonstrate histograms for the five positions for a latent factor of the player mode in a regular season. Each histogram reveals the frequency distribution of the normalized posterior means.